

OM of: US-09-303-518d-651 to: SPTRMBL_19:* out_format : pfs

Date: Jun 30, 2002 8:25 AM

About: Results were produced by the GenCore software, version 4.5.
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=frame+gfp.model -DEV=xlh
-Q/cg2_1/USPTO_SPOOL/US09303518/runat_28062002.142714.4338/app_query.fasta_1.23501
-DB=SPTRMBL_19 -QPMT=fastan -SUFFIX=tspt -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000
-GAPOP=4.500 -GAPEXT=0.050 -GAPOP=10.000 -XGAPEXT=0.500
-GAPOP=6.000 -GAPEXT=7.000 -GAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=100 -DLOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=45 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09303518 -CGCN1_1.1967 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPAY -WAIT -THREADS=1

Search information block:

Query: US-09-303-518D-651

Query length: 4350

Database: SPTRMBL_19.*

Database sequences: 562222

Database length: 172994929

Search time (sec): 883.700000

score_list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
sp_bacteria:051169	+ 1551.50	1630.54	1.1e-83	1561	! Q51169 neisseria meningitidis
sp_bacteria:09AE76	+ 1523.00	1600.37	5.5e-82	1552	! Q9AIU6 neisseria meningitidis
sp_bacteria:09AUXK3	+ 1522.00	1600.05	6.2e-82	1431	! Q9JXK3 neisseria meningitidis
sp_bacteria:093N34	+ 1452.00	1523.93	8.7e-78	1764	! Q9JX34 haemophilus aegyptius
sp_bacteria:09AUVB9	+ 1410.00	1479.34	2.6e-75	1773	! Q9JVB9 neisseria meningitidis
sp_bacteria:09K0B4	+ 1404.00	1472.77	6.0e-75	1815	! Q9K0B4 neisseria meningitidis
sp_bacteria:09S6X3	+ 1313.50	1382.26	1.2e-69	993	! Q9S6X3 neisseria meningitidis
sp_bacteria:09S6X2	+ 1313.50	1382.23	1.2e-69	997	! Q9S6X2 neisseria meningitidis
sp_bacteria:057309	+ 1310.00	1378.52	1.9e-69	996	! Q57309 neisseria meningitidis
sp_bacteria:057035	+ 1309.00	1377.50	2.2e-69	992	! Q50755 neisseria meningitidis
sp_bacteria:030575	+ 1308.50	1376.92	2.4e-69	997	! Q30575 neisseria meningitidis
sp_bacteria:030574	+ 1302.00	1370.04	5.8e-68	996	! Q9S6X4 neisseria meningitidis
sp_bacteria:09S6X4	+ 1285.00	1352.05	5.8e-68	992	! Q30573 neisseria meningitidis
sp_bacteria:030573	+ 1284.00	1350.99	6.7e-68	992	! Q9S6X5 neisseria meningitidis
sp_bacteria:09S6X5	+ 1281.00	1347.81	1.0e-67	992	! Q47692 escherichia coli. tsh.
sp_bacteria:047692	+ 1084.50	1136.48	4.3e-56	1377	! Q88093 escherichia coli. haem
sp_bacteria:088093	+ 1083.50	1135.42	4.9e-56	1372	! Q9AI58 shigella flexneri. 2a. p
sp_bacteria:054151	+ 1051.50	1101.52	3.8e-54	1373	! Q69746 neisseria lactamica. lga
sp_bacteria:09AL58	+ 1051.50	1101.51	3.8e-54	1373	! Q99G66 shigella flexneri. sepa
sp_bacteria:069746	+ 1036.00	1102.88	2.3e-53	192	! Q9AI58 shigella flexneri. 2a. p
sp_bacteria:069747	+ 1028.50	1094.83	6.3e-53	194	! Q69747 neisseria lactamica. lga
sp_bacteria:0990C6	+ 1003.00	1050.14	2.8e-51	1364	! Q99G66 shigella flexneri. sepa
sp_bacteria:054165	+ 1003.00	1050.13	2.8e-51	1366	! Q54165 shigella flexneri. sepa
sp_bacteria:09LA58	+ 995.00	1041.85	8.2e-51	1335	! Q9AI58 shigella flexneri. 2a. p
sp_bacteria:09LA54	+ 991.00	1037.61	1.4e-50	1335	! Q9AI54 escherichia coli. eaa
sp_bacteria:093AF7	+ 951.00	995.03	3.3e-48	1359	! Q93AF7 escherichia coli. autot
sp_bacteria:093SE3	+ 936.00	979.10	2.5e-47	1363	! Q93SE3 escherichia coli. hypot
sp_bacteria:051171	+ 870.00	918.05	1.7e-43	508	! Q51171 neisseria meningitidis
sp_bacteria:051167	+ 862.00	909.57	5.0e-43	508	! Q51167 neisseria meningitidis
sp_bacteria:051168	+ 855.00	902.14	1.3e-42	508	! Q51168 neisseria meningitidis
sp_bacteria:051165	+ 853.50	900.77	1.6e-42	496	! Q51165 neisseria meningitidis
sp_bacteria:051163	+ 848.00	894.94	3.4e-42	496	! Q51163 neisseria meningitidis
sp_bacteria:051173	+ 847.50	894.21	3.6e-42	507	! Q51173 neisseria meningitidis
sp_bacteria:051164	+ 844.00	890.48	5.8e-42	508	! Q51164 neisseria meningitidis
sp_bacteria:051172	+ 840.00	886.45	1.0e-41	496	! Q51172 neisseria meningitidis
sp_bacteria:051166	+ 834.00	880.09	2.3e-41	496	! Q51166 neisseria meningitidis
sp_bacteria:057386	+ 823.00	868.43	1.0e-40	496	! Q57386 neisseria meningitidis
sp_bacteria:051170	+ 815.00	859.94	3.0e-40	496	! Q51170 neisseria meningitidis
sp_bacteria:048205	+ 760.00	802.27	5.2e-37	462	! Q48205 haemophilus influenzae
sp_bacteria:069745	+ 733.00	781.68	1.8e-35	190	! Q69745 neisseria lactamica. lga

sp_bacteria:048204	+	687.00	724.68	1.1e-32	471	! Q48204 haemophilus influenzae
sp_bacteria:09E2E7	+	657.50	684.18	7.0e-31	1305	! Q9E2E7 escherichia coli. e
sp_bacteria:048206	+	653.00	688.23	1.1e-30	492	! Q48206 haemophilus influenzae
sp_bacteria:048159	+	644.00	679.30	3.7e-30	460	! Q48159 haemophilus influenzae
sp_bacteria:077070	+	638.00	663.49	1.0e-29	1306	! P77070 escherichia coli. es
sp_bacteria:068900	+	632.00	657.21	2.2e-29	1295	! Q68900 escherichia coli. oig
sp_bacteria:032555	+	631.50	656.64	2.4e-29	1300	! Q32555 escherichia coli. oig
sp_bacteria:032591	+	626.50	651.34	4.8e-29	1300	! Q32591 escherichia coli. pr
sp_bacteria:09F6T1	+	613.00	637.06	3.0e-28	1295	! Q9F6T1 escherichia coli. se
sp_bacteria:09L8L1	+	563.00	584.11	2.7e-25	1285	! Q9L8L1 shigella flexneri. 2a
sp_bacteria:09X748	+	498.00	538.29	1.2e-21	100	! Q69748 neisseria lactamica.
sp_bacteria:09XD84	+	322.50	331.46	4.1e-11	989	! Q9XD84 escherichia coli. tit
sp_bacteria:097P71	+	298.50	291.76	1.4e-09	4776	! Q97P71 streptococcus pneum
sp_bacteria:048031	+	277.00	279.23	2.1e-08	1536	! Q48031 haemophilus influen
sp_bacteria:09RNI2	+	272.50	274.33	4.0e-08	1557	! Q9RNI2 haemophilus influen
sp_bacteria:09XCJ4	+	271.00	270.32	5.1e-08	2035	! Q9XCJ4 salmonella typhimur
sp_fungi:Q9C105	+	266.50	270.06	8.6e-08	1236	! Q9C105 schizosaccharomyces
sp_bacteria:09XCQ3	+	262.00	261.10	1.7e-07	1963	! Q9XCQ3 salmonella typhimur
sp_bacteria:09KX33	+	260.00	254.57	2.4e-07	3194	! Q9KX33 helicobacter pylori
sp_bacteria:09S6R3	+	257.00	263.42	2.5e-07	3381	! Q9S6R3 streptococcus crista
sp_bacteria:09X4W2	+	256.00	247.14	4.0e-07	846	! Q9S6R3 escherichia coli. esp
sp_bacteria:09KSI2	+	255.00	246.05	4.4e-07	4545	! Q9X4W2 vibrio cholerae. rty
sp_bacteria:099QY4	+	252.00	249.18	5.1e-07	4558	! Q9KSI2 vibrio cholerae. rty
sp_bacteria:048028	+	251.00	252.01	6.8e-07	2271	! Q99QY4 staphylococcus aure
sp_bacteria:09F292	+	250.50	251.76	7.3e-07	1477	! Q48028 haemophilus influenza
sp_bacteria:0939N5	+	250.00	244.32	7.8e-07	1432	! Q9F292 yersinia pestis. yaf
sp_bacteria:093TY6	+	248.00	244.92	9.4e-07	3072	! Q939N5 streptococcus gordon
sp_bacteria:09F289	+	246.00	246.83	1.2e-06	2276	! Q93TY6 staphylococcus aure
sp_bacteria:09F289	+	246.00	238.39	1.4e-06	1457	! Q9F289 yersinia pestis. yaf
sp_bacteria:09PHN8	+	245.50	249.91	1.4e-06	3705	! Q9F289 yersinia pestis. yaf
sp_bacteria:09JMS3	+	244.50	245.79	1.7e-06	978	! Q9PHN8 campylobacter jejuni
sp_human:060585	+	242.00	247.79	2.3e-06	820	! Q9JMS3 escherichia coli. y
sp_bacteria:09JMS5	+	234.50	232.94	7.1e-06	1758	! Q60585 homo sapiens (human)
sp_bacteria:09PZU9	+	234.00	238.19	6.9e-06	928	! Q9PZU9 campylobacter jejuni
sp_bacteria:098JH8	+	233.50	235.24	7.7e-06	1213	! Q98JH8 rhizobium loti (meso
sp_bacteria:09ZD91	+	233.50	239.29	8.5e-06	2340	! Q9ZD91 rickettsia prowazek
sp_bacteria:09S6M9	+	232.00	236.30	9.0e-06	905	! Q9S6M9 bordetella pertussis
sp_invertebrate:076602	+	231.50	232.66	1.0e-05	1275	! Q9S3M8 bordetella pertussis
sp_bacteria:09ZKS9	+	230.50	225.88	1.3e-05	2399	! Q76602 caenorhabditis ele
sp_bacteria:025579	+	230.00	224.88	1.4e-05	2529	! Q9ZKS9 helicobacter pylori
sp_bacteria:089259	+	229.50	233.60	1.3e-05	910	! Q89259 helicobacter pylori
sp_bacteria:088143	+	227.00	230.90	1.8e-05	915	! Q88143 bordetella pertussis
sp_invertebrate:p91365	+	227.00	222.82	2.0e-05	2232	! P91365 caenorhabditis ele
sp_bacteria:09S6N0	+	226.00	229.93	2.0e-05	905	! Q9S6N0 bordetella pertussis
sp_bacteria:09L4E2	+	226.00	229.97	2.0e-05	911	! Q9L4E2 bordetella pertussis
sp_bacteria:069257	+	225.50	229.38	2.2e-05	907	! Q69257 bordetella pertussis
sp_bacteria:098E20	+	225.00	215.58	2.9e-05	3930	! Q98E20 rhizobium loti (meso
sp_bacteria:093L98	+	224.50	228.19	2.5e-05	920	! Q93L98 bordetella pertussis
sp_bacteria:09KWR3	+	224.50	220.40	2.9e-05	2178	! Q9KWR3 streptococcus gordon
sp_bacteria:09S6N1	+	223.50	227.23	2.8e-05	910	! Q9S6N1 bordetella pertussis
sp_bacteria:09X6I7	+	223.50	226.03	2.9e-05	1039	! Q9X6I7 escherichia coli. ar
sp_bacteria:09XC47	+	223.00	219.11	3.5e-05	2106	! Q9XC47 rickettsia australis
sp_rodent:070495	+	221.50	225.24	3.7e-05	897	! Q70495 mus musculus (mouse)

seq_name: sp_bacteria:051169

seq_documentation_block:

ID Q51169 PRELIMINARY; PRT; 1561 AA.

AC Q51169;

DT 01-NOV-1996 (TEMBLrel. 01, Created)

DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)

DT 01-DEC-2001 (TEMBLrel. 19, Last annotation update)

DE ICAL PROTEASE.

GN IGA.

OS Neisseria meningitidis.

OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.

OX NCBI_TaxID=487;

RP SEQUENCE FROM N.A.

RC STRAIN=HF13;

RX MEDLINE=95302961; PubMed=7783620;

RA Lomholt H., Poulsen K., Mogens K.;
 RT "Comparative characterization of the iga gene encoding IgA1 protease
 RT in Neisseria meningitidis, Neisseria gonorrhoeae and Haemophilus
 RT influenzae";
 RL Mol. Microbiol. 15:495-506(1995).
 DR EMBL: X82474; CAA57857.1; -;
 DR MEROPS; S06.001; -;
 DR InterPro: IPR002195; Dihydroorotase.
 DR InterPro: IPR000710; Iga_S6.
 DR Pfam: PF02395; IGA1; 1.
 DR PRINTS; PR00921; IGASERPTASE.
 DR PROSITE; PS00482; DIHYDROOROTASE_1; UNKNOWN_1.
 KW Protease.
 SQ SEQUENCE 1561 AA; 171849 MW; 1C96E291A00017D5 CRC64;

alignment_scores:
 Quality: 1551.50 Length: 1754
 Ratio: 1.730 Gaps: 51
 Percent Similarity: 51.140 Percent Identity: 26.910

alignment_block:

US-09-303-518d-651 x Q51169 ..

Align seg 1/1 to: Q51169 from: 1 to: 1561

```

49 AAACCGCGCGCATCCGCTCTCTCGCTGCTACTAGCATATGCGTGC 98
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2 LysThrLysArgPheLysIleAsnAlaIleSerLeuSerIlePheLeuAl 18
99 GTTCGGCATCTCTCCCAAGCTTGGCGGGGACACACTATTTCGGCATCA 148
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
18 atyAlaLeuThrProTySerGluAlaAlaLeuValArgAspValA 35
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
149 ACTACCACTACTCGCACTTTCGCGAAATAAGCAAGTTTTCAGTGC 198
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
35 spfyGlnIlePheArgPheAlaGluAsnLysGlyLysPhePheVal 51
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
199 GGGGGAAGATATTAGTGTACAAACAAAGGGAGTTGGTCGGCAA 248
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
52 GlyAlaThrAspLeuSerValLysAsnLysGlnGlyGlnAsnIleGlyAs 68
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
249 ATCATGCAAAAGCCGATGATTGATTTTCTGCTGCTGCGCTAAC. 297
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
68 nAlaLeuSerAsnValProMetIleAspPheSerValAlaAspValAsnL 85
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
298 ..GGCGTGGCGCATGTGGCGCATCAATATGTCAGCGTGGCATAC 345
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
85 ysArgIleAlaThrValValAspProGlnTyAlaValSerValLysHis 101
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
346 .....AACGGCGCTATAACACGTTGA 368
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
102 AlaLysAlaGluValHisThrPheTyTyTyGlyGlnTyAsnGlyHisAs 118
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
369 TTTTGTGGGAGGAGGAGNAATCCCGATCAGCACCGCTTTTCTTACCAA 418
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
118 nAspValAlaAspLysGluAsn.....GluTyArgV 129
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
419 TTGTGAAAAGAATAATTATAAGCTGACAAATTCACACCTTTACACCGC 468
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
129 alValGluGlnAsnAsnTyGluProHisLysAlaTrpSerAlaSerAsn 145
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
469 .....GATTANCATATGCGCGCTTTCATATAATTGTCTAC 503
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
146 LeuGlyArgLeuGluAspTyAsnMetAlaArgPheAsnLysPheValTh 162
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
504 AGATGCAAGACCTGTGCAATACACAGTGCATATGAGGGG.....AATAC 550
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
162 rGluValAlaProIleAlaProThrAspAlaGlyGlyLeuAspThrT 179
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
551 ATTCGGATAAAGAAAATATCCGCGGTGTCGGCATCGCTCGCAGACAC 600
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
179 yrlLysAspLysAsnArgPheSerPheValArgValGlyAlaGlyArg 195
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

601 CACTAT.....TGCGCTTATGATGATGACAAACACGG 632
||| ||||| ||| ||||| ||| ||||| |||
196 GlnLeuValTyGluLysGlyAlaTyHisGlnGluGlyAsnGluLysG 212
||| ||||| ||| ||||| ||| ||||| |||
633 C.....GATTATCTACTCCGCGCATGTTAATTGGCGG. 669
| ||||| ||| ||||| ||| ||||| |||
212 yTyAspLeuArgAspLeuSerGluAlaTyArgTyAlaIleAlaGlyT 229
||| ||||| ||| ||||| ||| ||||| |||
670 .....AATACACAT...ATG 681
||||| ||||| ||||| ||||| ||||| ||||| |||||
229 hrProTyLysAspIleAsnIleAspGlnThrMetAsnThrGluGlyLeu 245
||| ||||| ||| ||||| ||| ||||| |||
682 CAGGGTTGGGAAATAATGCGTANTTAGTTGACGCGCATGTGCGCA 731
||| ||||| ||| ||||| ||| ||||| |||
246 IleGlyPheGlyAsnHisAsnThrHisTySerAlaGluGluLeuLysG 262
||| ||||| ||| ||||| ||| ||||| |||
732 TGCCAACGACTATGGCCCTATCCGATTGCGGCGGCGACAGCG 781
||||| ||||| ||||| ||||| ||||| ||||| |||||
262 nAlaLeuSerGlnAspAlaLeuThrAsnTyGlyValLeuGlyAspSerG 279
||| ||||| ||| ||||| ||| ||||| |||
782 GTTCGCCAATGTTTATTATGACAAACAAACAATAATGGTCTCAAC 831
||||| ||||| ||| ||||| ||| ||||| |||
279 lySerProLeuPheAlaPheAspLysGlnLysAsnGlnTrpValPheLeu 295
||| ||||| ||| ||||| ||| ||||| |||
832 GGAGTTTACAAACCGCTACCTTATTCGGCAGGAAACGGTTTCCA 881
||||| ||||| ||||| ||||| ||||| ||||| |||||
296 GlyThrTyAspTyTrpAlaGlyTyGly..... 305
||||| ||||| ||||| ||||| ||||| ||||| |||||
882 GCTGATACGCAAGATTGGTTCTACGATGACATTTACAGAGGCGATACAC 931
||||| ||||| ||| ||||| ||| ||||| |||
306 .....LysLysSerTrpGlnLysTrpAsnLysLysGlu...P 319
||||| ||||| ||| ||||| ||| ||||| |||
932 ATACCGTCTNTTTTGAACCGCGCAGTAACGGA..... 963
||| ||||| ||| ||||| ||| ||||| |||
319 heAlaAspLysIleLysGlnArgAspAlaGlyThrIleLysGlyTy 335
||| ||||| ||| ||||| ||| ||||| |||
964 .....CATTTTCCTTACATCCAAACACGCTACGGGTACGCT 1004
||||| ||||| ||| ||||| ||| ||||| |||
336 GlyGluHisIleTrpLysThrGlyThrAsnSerHisIleGlySerTh 352
||||| ||||| ||| ||||| ||| ||||| |||
1005 AACAGAAACCAAGAAAGTTCCTCAATCCAAAGCTTAAAGTACAGACAG 1054
||| ||||| ||| ||||| ||| ||||| |||
352 r.....Alav 354
||||| ||||| ||||| ||||| ||||| ||||| |||||
1055 TCGACTGTTTGCAATCTTGAATGAACACTATTAAGAACAGTTTAC 1104
||||| ||||| ||| ||||| ||| ||||| |||
354 alArgLeuAlaAsn.....AsnGluArgAsp..... 362
||||| ||||| ||||| ||||| ||||| ||||| |||||
1105 GCGCAGGGGTGTTAATCAGTACCGTCCAAAGTTAAACACACGCTGAAA 1154
||||| ||||| ||||| ||||| ||||| ||||| |||||
363 .....AlaAsnAsnGlyGlnAs 368
||||| ||||| ||||| ||||| ||||| ||||| |||||
1155 CTTTCTTTTATCGATTACGCAACGCAACACTCATCTTATCAACAACA 1204
||||| ||||| ||| ||||| ||| ||||| |||
368 nValThrPhe.....GluAsnAsnGlyThrLeuValLeuAspGlnAsnI 383
||||| ||||| ||| ||||| ||| ||||| |||
1205 TCACCAAGCGCGGGGTTTGTATTTTGAAGTGTATTTTACGGTCTCG 1254
||||| ||||| ||| ||||| ||| ||||| |||
383 leAsnGlnGlyAlaGlyLeuPhePheLysGlyAspTyThrValLys 399
||||| ||||| ||| ||||| ||| ||||| |||
1255 CCTGAAAACCAACGAA...ACGTGGCAAGCGCGGCTTCATATCAGTGA 1301
||||| ||||| ||||| ||||| ||||| ||||| |||||
400 GlyAlaAsnAsnAspIleThrTrpLeuGlyAlaGlyIleAspValAlaAs 416
||||| ||||| ||| ||||| ||| ||||| |||
1302 AGACAGTACCGTTACTTGGAAAGTAAACGCGGTGGCAACAGCGCTGT 1351
||| ||||| ||| ||||| ||| ||||| |||
416 pGlyLysLysValValTrpGlnValLysAsnProAsnGlyAspArgLeuA 433
||||| ||||| ||| ||||| ||| ||||| |||
1352 CCAAAATCGGCAAGGACGCTGCACGTTCAAGCCAAAGGGGAAACCAA 1401
||||| ||||| ||| ||||| ||| ||||| |||
433 laLysIleGlyLysGlyThrLeuGluIleAsnGlyThrGlyValAsnGln 449
||||| ||||| ||||| ||||| ||||| ||||| |||||

```

1402 GGCTCGATACGCTGGCGACGGTACAGTCACTTTGGATCAGCAGCGAGA 1451
 450 GlyLeuLeuLysValGlyAspGlyThrValIleLeuAsnGlnLysAlaAs 456
 1452 CGATAAAGGCAAAAACAGCCTTTAGTGAATACGCTTGNTCAGCGCA 1501
 466 pSerAsnGlnLysValSerAlaPheSerGlnValcIyIleValSerGlyA 483
 1502 GGGGTACGGTGCACATGAATCCGATACCGGATATCACTAGTTCAACCCGACAAACTC 1551
 483 rgGlyThrLeuValLeuAsnSerSerAsnGlnIleAsnProAspAsnLeu 499
 1552 TATTTCCGCTTTCGGCGCGAGCTTGATTTAAACGGCGATTCGCTTTC 1601
 500 TyrPheGlyPheArgcIyGlyArgLeuAspAlaAsnGlyAsnAspLeuTh 516
 1602 GTTCCACCGTATTCAAAATACCGATGAAGGGCGCATGATTGNCNATCATA 1651
 516 rPheGluHisIleAlaGsnValAspGluGlyAlaArgIleValAsnHisA 533
 1652 ATGCCACAACATCCACCGTTACCATTCACAGGAATGAAGATATTACA 1701
 533 snThrSerHisAlaSerThrIleThrLeuThrGlyLysSerLeuIleThr 549
 1702 CAACCGAGT..... 1710
 550 AsnProAsnSerLeuSerValHisSerIleGlnAsnAspTyrAspGluAs 566
 1711GGTAAGA 1717
 566 pAspTyrSerTyrTyrTyrArgProArgArgProIleProGlnGlyLysA 583
 1718 ATATCAATAGACTTAATTAC..... 1737
 583 spLeuTyrTyrLysAsnTyrArgTyrTyrAlaLeuLysSerGlyGlySer 599
 1737 1737
 600 ValAsnAlaProMetProGluAsnGlyGlnThrGluAsnAsnAspTrpI 616
 1738ACCAAGAAATTCGCTACAAC.... 1758
 616 eLeuMetGlySerThrGlnGluGluAlaLysLysAsnAlaMetAsnHisL 633
 1759GTTGGTTGGCGAGAAGAT 1779
 633 ysAsnAsnGlnArgIleSerGlyPheSerGlyPhePheGlyGluGluAsn 649
 1780 ACGACCAAAACGACGGCGCTCAACTGTTTTACAGCCCGCCGCGAGA 1829
 650 GlyLysGlyHisAsnGlyAlaLeuAsnLeuAsnPheAsnGlyLysSerAl 666
 1830 AGACCGCACCCNGCTGTCTTCGGCGGAACAAATTTAAACGGCAACATCA 1879
 666 agIleAsnArgPheLeuLeuThrGlyGlyThrAsnLeuAsnGlyLysIleS 683
 1880 CGCAACAAACGGCAAACTGTTTTTCAGCGGACAGCAGACCGCACGCGC 1929
 683 erValThrGlnGlyAsnValLeuLeuSerGlyArgProThrProHisAla 699
 1930 TACAATCATTTAGGAAGCGGTGTGCTAAAAATCGAAGGT.....ATCCC 1973
 700 ArgAspPheValAsnLysSerSerAlaArgLysAspAlaArgPheSerLy 716
 1974 ACAAGGAGAAATCTGTGGGACAAACGACTGGATTCNACCGCATTTAAAG 2023
 716 sasAsnGluValValPheGluAspAspTrpIleAsnArgThrPheLysA 733
 2024 CGSAAATTTCCATATT...CAGGCGGCGACGCGGTGATTTCCGCGCAAT 2070
 733 laAlaGluIleAlaValAsnGlnSerAlaSerPheSerSerGlyArgAsn 749
 2071 GTTGCCAAAGTGAAGCGGATTGNCATTTGACCAATTCAGCCGCAAGCAGT 2120

750	ValSerAsnIleThrAlaAsnIleThrAlaThrAspAsnAlaLysValAs	766
2121	TTTGTGTGTCGCACCGCATCAAGAGCATACATCTGTACACGTTCCGGACT	2170
766	nLeuGlyTyr.....LysAsnGlyAspLueValcysValargSerAsp	781
2171	GGACNGGTCTGACAAATCTGTCTGAANAANCATTTACCGAGATAAAGTG	2220
781	yrThrGlyTyrValThrCys.....	787
2221	ATTGCTTCATTGACTAAGACNGACNTNAGCGGCANTGTNAGCTNNCCNA	2270
788AsnThrAsp.....	790
2271	TNACGNTNNTTNAANCTCNCNGGCNTGCNNCACTNNAANGCAATCTTA	2320
791AsnLeus	793
2321	GTGCAAAATGGCGATACAGTTATACAGTCAGCCACACAGCCACCCAA...	2367
793	er.....AspLysAlaLeuAsnSerPheAspAlaThrGlnIle	805
2368	AACGGCAACCTTAGGCTCTGTGGCAATGCCCAAGCAACATTTAATCAAGC	2417
806	AsnGlyAsnValAsnLeuAsnGlnAsnAlaAlaLeuValLeuGlyLysAl	822
2418	CACATTAAAGCGACACNCATCGGNTCTGGGCAATGCTTCATTAACTAA	2467
822	aAlaLeuTrpGlyGlnIleGlnGlyGlnGlyAsnSerSerValSerLeuA	839
2468	GCAACACCGCCACAAAAAGCGCAGCTGTGACGCTTCCGCACACGCTAAG	2517
839	snGlnHis.....	841
2518	GCAACGTTAAGGCATTCCGCACCTCAAGGCAATCTCTCCCTACGCCGATA	2567
841	841
2568	GGCAGTATTCCATTTTGAAGACAGCGCTTTACCGGACACTCAGCGGCA	2617
841	841
2618	GCAAGGANNACGATTACACTTAAAGACAGACGAGTGAAGCGCTCCGCTCA	2667
842SerLysTrpHisLeuThrSer	848
2668	GGCAGGGAATTAGCAATTTAAACCTTCACACGCCACCATTTACACTCAA	2717
849	AspSerGlnValHisAsnLeuSerLeuAlaAspSerHisIleHisLeuAs	865
2718	TTCCGCCTATCGCCACGATGCTGACGGCGGCAACCGCAGNGTGTCAAG	2767
865	nAsnAla.....SerAspAlaGlnSerAla.....	873
2768	ACACGCCGCGCGCTTCGCGCGCTTCCCTATATATCCGTTACACGCCCA	2817
873	873
2818	ACTTCGGTAGAATCCCGTTTCAACACGCTGACGATAAAGCGCAAAATTGAA	2867
874AsnLysTyrHisThrLeuLysIleAsn.....HisLeuSe	885
2868	CNGTCAAGCAACATTCGCTTTATGTGGAACTCTTCGGGTACCGAAGCG	2917
885	rGlyAsnGlyHisPheHisTyrLeuThrHisLeuAlaLysAsnLeuGlyA	902
2918	ACAAATTCGAAGCTGGCGAAGTTCCGAAGNACTTACACCTTGGCGGTC	2967
902	sPlsValLeuValLysGluSerAlaSerGlyHisTyrGlnLeuHisVal	918
2968	ACAATATCCGCGACACGAACCC.....GTAAGCTCGATCAATTT	3005

[illegible]

783 TTCGCCAATGTTATTATGACAAACAACAATAAATGGCTGCTCAAGC 832
|||||::||| |||||::||| ::::::|||::|||::|||::|||::|||
268 ySerProLeuPheAlaTyAspLysGlnGlulYsArgTtpValPheLeuG 285
ySerProLeuPheAlaTyAspLysGlnGlulYsArgTtpValPheLeuG 285
333 GAGTTTTACAA.....ACCGGTCACCTTATTCCGCACGGGAACGGT 876
|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
285 lYAlaTyAspTyTrpAlaGlyTyr.....GlnLYsAsnSer 297
lYAlaTyAspTyTrpAlaGlyTyr.....GlnLYsAsnSer 297
877 TTCCAG.....CTGATACCAGAAGATTGGTGTTCTACGATGACATT...TA 917
:::|||::||| |||| | ||| ||| ||| ||| ||| ||| ||| ||| |||
298 TrpGlnGlutrpAsnIleTyrlYsLysGluPheAlaAspLysIleargGI 314
TrpGlnGlutrpAsnIleTyrlYsLysGluPheAlaAspLysIleargGI 314
918 CAGAGGGCATACACATACCGTCTNTTTTGAAACCGCGCACGTAAACGGA.... 963
||| ::::|||::|||::|||::|||::|||::|||::|||::|||::|||
314 nArgAspAsnAlaGlyThrIle.....LYsGLyAsnGlyCluh 327
nArgAspAsnAlaGlyThrIle.....LYsGLyAsnGlyCluh 327
964 ..CATTTTTCTTACATCCAACAACAACACCGTAGCGGTACAGCAA 1011
|||||::|||::|||::|||::|||::|||::|||::|||::|||::|||
327 iSHisTrpAsnIleThrPheGly..... 334
iSHisTrpAsnIleThrPheGly..... 334
1012 ACCAACGAAAAGGNTCCAAATCCAAAGCTTAAAGTACAGACAGTCGCGACT 1061
|||||::|||::|||::|||::|||::|||::|||::|||::|||::|||
335 ThrAsnSerLYsIleGlySer.....ThralaValargLe 346
ThrAsnSerLYsIleGlySer.....ThralaValargLe 346
1062 GTTTGACGAATCTTTGTGAAGAACTGATAAAGAACCACTTTACGCGCAG 1111
| ::|||::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
346 u.....AlaGlyAsnGluLYsasp..... 352
u.....AlaGlyAsnGluLYsasp..... 352
1112 GGGGTGTTATTCAGTACCGTCCAAGGTTAAACAACGGTGAACACTTTCT 1161
|||||::|||::|||::|||::|||::|||::|||::|||::|||::|||
353AlaAsnGlnGlyInAsnValThr 360
.....AlaAsnGlnGlyInAsnValThr 360

858SerAspAlaGlnSerAla..... 863
2775 CGCGCGCGTTCGCGCGTTCCTTATTCCTGTTACACGCCCACTTCGG 2824
863 863
2825 TAGAATCCGCTTCAACACGCTGACGTAACGCAAAATTAACNGTCAA 2874
864AsnGlnTyrHisThrLeuLysIleAsn...HisLeuSerGlyAsn 877
2875 GGAACATCCGCTTATGTCGGAACCTTCGCTACCGAAGCAGCAAAAT 2924
878 GlyHisPheHisTyrLeuThrHisLeuAlaGluAsnLeuGlyAspLysVa 894
2925 GAAGCTGGCGAAGTTCGGAAGNACTTACACCTTGGCGGTCAACAATA 2974
894 iLeuValLysGluSerAlaSerGlyHisTyrGlnLeuHisValGlnAspL 911
2975 CCGCAACGACCC.....GTAAGCCTCGATCAATTGACGGTA 3012
911 ystThrGlyGluProAsnGlnGluGlyLeuAsnLeuPheAspAlaSerSer 927
3013 GTGAAGGGAAGACACAAACCGCTGTCGAAACCTTAATTTACCCCT 3062
928 ValGlnAspArg.....SerArgLeuSerValSerLe 938
3063 GCAAAACGACACGCTGATCCGCGCGTTCGGAACCTTAATTTACCCCT 3112
938 uAlaAsnAsnHisValAspLeuGlyAlaLeuArgTyrThrIleLysThrG 955
3113 AAGACGGCGAGTTCGCGCTGATAATCCGCTCAAGAA..... 3150
955 LuAsnGlyIleThrArgLeuTyrAsnProTyrAlaGluAsnArgArgArg 971
3150 3150
972 ValLysProValProSerProAlaThrAsnThrAlaSerGlnAlaGlnTh 988
3150 3150
988 rAspSerAlaGlnIleAlaLysProGlnAsnIleValValAlaProProS 1005
3151CAAGAGCTTTCGACAACTCGCAAGGCA 3180
1005 erProGlnAlaAsnGlnAlaGluAlaLysArgGlnGlnAlaLysAla 1021
3181 GAAGCCAAATA.....CAGCGGAAACACACGCGCA... 3216
1022 GluGlnValLysArgGlnGlnAlaGluAlaGluArgLysSerAlaGluLe 1038
3217AGCCTTCACGCGCTGATTGCGCGCGGC 3244
1038 uAlaLysGlnLysAlaGluAlaGluArgGluAlaArgGluLeuAlaThra 1055
3245 GCATCCCGCGGAAAGACAGAGCGTTGCGCAACCGCGCGG..... 3288
1055 rGlnLysAlaGluGlnGluArgSerSerAlaGluLeuAlaArgHis 1071
3288 3288
1072 GluLysGluArgGluAlaAlaGluLeuSerAlaLysGlnLysValGluAl 1088
3288 3288
1088 aGluArgGluAlaGlnAlaLeuAlaValArgLysAlaGluAlaGluG 1105
3289CNG 3291
1105 LuAlaLysArgGlnAlaAlaGluLeuAlaArgGlnGlnGluAlaArg 1121
3292 GCAGCGGGGAAATGTCGGCTATTGCAAGCGGAGGAGAGAGAAAA... 3339
..... 3339

1122 LysAlaAlaGluLeuAlaLysGlnLysAlaGluThrGluArgLysAl 1138
3340CGGGTCAGCGCGGTAAGACACGCGN...T 3367
1138 aAlaGluIleAlaGluGlnLysAlaGluArgGluAlaGluL 1155
3368 TGGCGAAACAGCGC...GAAGCGGAACCCGCGNTACCAACGCCCTTC 3414
1155 euAlaLysGlnLysAlaGluGluGlyArgGlnAlaGlnSerGln 1171
3415 CCGCGCGCGCGCGCGC..... 3435
1172 ProLysArgArgAsnArgAlaIleProGluLeuSerSerAspAl 1188
3436CGGGATTTCGCG..... 3447
1188 aThrThrArgAlaLeuProArgIleAlaArgAsnSerAsnProAspAlaS 1205
3447 3447
1205 erAspTyrGluLysIleProLeuAspAlaLeuGluAspGluAspValSer 1221
3448CAACGCGCAG..... 3456
1222 GluSerValAspThrSerAspLysGlnProGlnAspAsnThrGluLeuHi 1238
3457CCCCAACCGCAAC 3469
1238 sGluLysValGluThrValSerLeuGlnProArgAlaAlaGlnProArgA 1255
3470 CTCACCCCAACCGCGCGACCTGATGAGCGGTATGCGCAATAGCGGT 3519
1255 iGlnAlaAlaGlnAlaAspAlaValSerThrAsnThrAsnSerAla 1271
3520 TTGAGTGAATTTTCGCGCACGCTCAACAGCGTTTCGCC...GTACAGGA 3566
1272 LeuSerAspAlaMetAlaSerThrGlnSerIleLeuLeuAspThrGlyAl 1288
3567 GAATTGACCGCGTTCGCGAAGACCGCGC.....A 3601
1288 aSerLeuThrArgHisIleAlaGlnLysSerArgAlaAlaGluLysA 1305
3602 ACGCGTTTGGACAGCNGCATCCGNNACCAACACTACCGTTCGCA 3651
1305 snSerValTrpMetSerAsnIleGlyTyrGlyArgAspTyrAlaSerAla 1321
3652 GATTTCCGCGCTACCGCCAAACACCGACCTCGCCAAATCGTATGCA 3701
1322 GlnTyrArgArgPheSerSerLysArgThrGlnThrGlnIleGlyIleAs 1338
3702 GAAACCTCGCACGCGG...CGCGTCGCGATCCTGTTTCGCACAAC 3748
1338 pArgSerLeuSerGluAsnMetGlnIleGlyValLeuLeuThrTyrSerA 1355
3749 GGACGAAACACTTCGACACGCGCATCGCAACTCGGACGCGCTTC 3798
1355 spSerGlnHisThrPheAspGlnAlaSerGlyLysAsnThr...PheVal 1370
3799 CACGCGCGCGTTTCGCGCAATACGCGATCGC...AGGTCGACATCGG 3845
1371 GlnAlaAsnLeuTyrGlyLysTyrTyrLeuAsnAspAlaTrpTyrValAl 1387
3846 CATCAGCAGCGCGCGGTTTAGCAGCGCGCANTCTNTCAGACGCGATCG 3895
1387 aGlyAspIleGlyAlaGlySerLeuArgSerArgLeuGlnThrGlnGlnL 1404
3896 GAGCAAAATCCGCGCGCGTCTGCTGATACGGCATTCAGGCACGATAC 3945
1404 yAlaAsnPheAsnArgThrSerIleGlnThrGlyLeuLeuGlyAsn 1420
3946 CGCGCGGTTTCGCGGATTCGCGCATCGAACCTACATCGCGCGCACCGG 3995
1421 ThrLeuLysIleAsnGlnPheGluIleValProSerAlaGlyIleArgTy 1437

```
3996 CTATTTCGTCACAAAGCGGATACCGCTACGAAACGTCACATATCGCCA 4045
      :::::::::::::::::::: ||| ::::::::::
1437 rSerArgLeuSerSerAlaAspTyrLysLeuGlyAsnAspSerValLysV 1454
      :::::::::::::::::::: ||| ::::::::::
4046 CCCCGGCTTCGCTTCAACCGNTACGNGCGCATTAAGCGAGATTAT 4095
      :::::::::::::::::::: ||| ::::::::::
1454 alSerSerMetSerValLysThrLeuThrAlaGlyLeuAspPheAlaTyr 1470
      :::::::::::::::::::: ||| ::::::::::
4096 TCATTCAACCGCGCGCAACACATNTCCATCNCCTTATTNAGCCCTGTC 4145
      ||||| :::::::::: ||| ::::::::::
1471 ArgPheLys...ValGlyAsnLeuThrValLysProLeuLeuSerAlaI 1486
      ||||| :::::::::: ||| ::::::::::
4146 CTAT...ACCGATCGCGTTCGCGCAACGTCGCAACACGCGTCAATACCG 4192
      :||| :::::::::: ||| ::::::::::
1486 atyPheAlaAsnTyrGlyLysGlyGlyValAsnValGlyLysAsnSerP 1503
      :::::::::::::::::::: ||| ::::::::::
4193 CNGTATTGGCTCAGGATTTCGCGCAACCGCGCATCGGAATGGCGGTA 4242
      :::::::::::::::::::: ||| ::::::::::
1503 heAlaTyrLysAlaAspAsnGlnGlnTyrSerAla.....GlyAla 1517
      :::::::::::::::::::: ||| ::::::::::
4243 AACGCCGAATCAAGTTTCACGCTGTCCTCCACGCTGCCGCGGCCAA 4292
      :::::::::::::::::::: ||| ::::::::::
1518 AlAlaLeuLeuTyrArgAsnValThrLeuAsnValAsnGlySerIleThrLy 1534
      ||||| :::::::::: ||| ::::::::::
4293 AGGNCGCCCACTGGAAGCGCAACACAGCGCGGCATCAATATTAGGCTACC 4342
      ||||| :::::::::: ||| ::::::::::
1534 scLysGlnLeuGluLysGlnLysSerGlyLleLysIleGlnIleA 1551
      ||||| :::::::::: ||| ::::::::::
4343 GCTGG 4347
      ||::
1551 rgPhe 1552
```

seq_name: sp_bacteriap:Q9JXK3

```
seq_documentation_block:
ID Q9JXK3 PRELIMINARY; PRT; 1431 AA.
AC Q9JXK3;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE SERINE-TYPE PEPTIDASE.
GN NMB1998.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MC58 / SEROGROUP B;
RX MEDLINE=20175755; PubMed=10710307;
RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mason T., Ciecko A., Parksey D.S., Blair E., Ciftone H., Clark E.B.,
RA Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J.,
RA Gill J., Scarlato V., Massignani V., Pizza M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
RA "Complete genome sequence of Neisseria meningitidis serogroup B strain
RT MC58."
RL Science 287:1809-1815 (2000).
DR EMBL; AE002549; AAF42325.1; -.
DR MEROPS; S06.0PW; -.
DR TIGR; NMB1998; -.
DR InterPro; IPR000710; IGA_S6.
DR Pfam; PF02395; IGA1; 1.
DR PRINTS; PR00021; IGASERPTASE.
KW Complete proteome.
SQ SEQUENCE 1431 AA; 157645 MW; 6801F0E209C24EFC CRC64;
```

alignment_scores:

Quality: 1522.00

Length: 1639

```
Ratio: 1.770 Gaps: 51
Percent Similarity: 52.471 Percent Identity: 27.639
alignment_block:
US-09-303-518d-651 x Q9JXK3
Align seg 1/1 to: Q9JXK3 from: 1 to: 1431
55 GGCGCGCATCGCTTCTCGCTTCTAGCCATATGCTCGTCTCGTGG 104
||| :::::::::::::::::::: ||| ::::::::::
17 GlyLeuPheAlaValSerProAlaTyrSerSerIleValArgAsn..... 31
||| :::::::::::::::::::: ||| ::::::::::
105 CATTCCTCCCAAGCTTGGCGGGACACACTATTTCGCATCAACTACC 154
::: :::::::::::::::::::: ::::::::::::::
32 .....AspValAspTyrG 36
155 AATACTATCGGACTTTCGCGAAATAAGCAAGTTTTCAGCTCGGGG 204
||||| :::::::::::::::::::: ||| ::::::::::
36 InTyrPheArgAspPheAlaGluAsnLysGlyAlaPheThrValGlyAla 52
||| :::::::::::::::::::: ||| ::::::::::
205 AAAGATATTGAGTNTACAAACAAAGGGAGTTTGGTCGCCAAATCAAT 254
::: :::::::::::::::::::: ||| ::::::::::
53 SerAsnIleSerIleGlnAspLysGlnGlyLysIleLeuGlyArgValle 69
255 GACAAAGCCCGCATGATTGATTTTCTGTGTGTCGCGT...AACGGCG 301
::: ||||| :::::::::: ||| ::::::::::
69 uAsnGlyIleProMetProAspPheArgValSerAsnArgGlnThrAlaI 86
::: ||||| :::::::::: ||| ::::::::::
302 TGCGCGCATTTGGCGGCGATCAATATATTGTGAGCGTGGCACATAACGC 351
::: ||||| :::::::::: ||| ::::::::::
86 leAlaThrLeuValHisProGlnTyrValAsnSerValLysHisAsnVal 102
::: ||||| :::::::::: ||| ::::::::::
352 GGCTATAACAACGTTGATTGTCGCGGAAGAGNAATCCCGATCAGCA 401
||||| :::::::::::::::::::: ||| ::::::::::
103 GlyTyrGlySerIleGlnPheGlyAsnAspThrGlnAsnProGluGluG 119
||| :::::::::::::::::::: ||| ::::::::::
402 CCGTTTTCTTACCAATTTGTGAAAGAAATATATATAGCCCTGCAAT 451
: :::::::::::::::::::: ||| ::::::::::
119 nAlaTyrThrTyrArgLeuValSerArg.....Asn 130
452 CACACCCCTTACAACGCGATTATCATATATGCGGTTTGCATATAATTTGTC 501
||||| :::::::::::::::::::: ||| ::::::::::
130 roHisPro...AspTyrAspTyrHisLeuProArgLeuAsnLysLeuVal 145
||| :::::::::::::::::::: ||| ::::::::::
502 ACAGATCAGAACCTGTGGAATGACGAGT..... 531
||| :::::::::::::::::::: ||| ::::::::::
146 ThrGluIleSerProThrAlaLeuSerSerValProLeuLeuGlyAsnG 162
532 .GACATGAGGGGAATACCTATTCCGATAAAGAAATAATATCCCGAGCGTG 580
::: :::::::::::::::::::: ||| ::::::::::
162 yGlnProLysAlaAsnAlaTyrLeuAspThrAspArgPheProTyrPheV 179
581 TCCGCATCGGCTCAGGACACCACTATTGGCGTTTATGATGATGACAAACAC 630
||| :::::::::::::::::::: ||| ::::::::::
179 alArgLeuGlySerGlyThrGlnGlnValArgLysAlaAspGlyThrArg 195
631 GCGGATTATCCCTACTCCCGCGCATGTTTAAATTTGGCGGCAAT...ACACA 677
::: :::::::::::::::::::: ||| ::::::::::
196 ThrArgThrAlaProAlaProAlaTyrGlnTyrLeuThrGlyGlyThrProLeuLy 212
678 TATCAGGTTTGGGAAATAATGGCGTANTTAGTTTGAGCGGCGATGTGC 727
::: ||||| :::::::::: ||| ::::::::::
212 sValLeuGlyPheGlnAsnHisGlyLeuLeu...ValGlyGlySerLeuT 228
728 GCCATGCCAACGACTATGGCCCTATGCGGATTCGAGTGGCGGCGGCGAC 777
||| :::::::::::::::::::: ||| ::::::::::
228 hrAspGln.....ProLeuAsnThrTyrAlaIleAlaGlyAsp 240
778 AGCGGTTTCGCAATGTTTATTATGACAAACAAACAAATAATATGCGTGT 827
||||| :::::::::::::::::::: ||| ::::::::::
241 SerGlySerProLeuPheAlaPheAspLysHisGluAsnArgTrpValLe 257
828 CAACGGAGTTTTCACAAACCGGCTACCCTTATTCGCGGAGGGAACGGTT 877
```


527 uAlaGlyValLeuSerThr.....TyrAlaGlyPheAsnAspPheP 271
528 TCCAG.....CTGATACGCAAGATTGGTCTACGATGACATTTACAGA 921
529 heAsnLysTyrIleValThrGlnProGluPheLeuArgSerThrIleArg 287
530 GGCATACACATACCGTCTTTTGAACCGCGC.....AG 956
531GlnTyrGluThrArgLeuAspValGlyLeuTh 298
532 TAACGGACATTTTCCTTTACATCCAAACACAGGTACGGGTACGATA 1006
533 rThrAsnGluLeuIleThrArgAspAsnGlyAsnGlyAsnSerThrLeuG 315
534 CAGAAACCAACGAAAG.....GTNTCCAATCCAAAGCTTAA 1044
535 lngLysLeuAsnGluArgIleThrLeuProIleAlaAsnProSerLeuAla 331
536 GTACAGACATCCGACTGTTTGAAGTCTTTGAATGAACCTGATAAAGA 1094
537 ProGln.....AsnAsp..... 335
538 ACCAGTTTACGGCGCGGGGGTGTATCATGATACCGTCCAGGTTAAACA 1144
539SerArgHisMetProSerGluAspA 344
540 ACAGTGAACCTTTCTTTTATCGAT.....TACGGCAACGCAAACTCATC 1191
541 laGlyLysThrLeuIleLeuSerSerArgPheAspAsnLysThrLeuMet 360
542 TTATCAACACATCAACCAAGCGGGGGTGTATTTGAAGGTGA 1241
543 LeuAlaAspAsnIleAsnGlnGlyAlaGlyAlaLeuGlnPheAspSerAs 377
544 TTTTACGGTCTCGCTGAAACCAACGAAACGTCGGGAGCGGGGGTTC 1291
545 nPheThrValValGlyLysAsnHis...ThrTrpGlnGlyAlaGlyValI 393
546 ATATCAGTGAAGCAGTACCGTCTTGTGAAGTAAACGCGTGGCAAC 1341
547 leValAlaAspGlyLysArgValPheTrpGlnValSerAsnProLysGly 409
548 GACCGCTGTCCAAATCGCAAGGACGCTGCAGCTTCAAGCCAAAG 1391
549 AspArgLeuSerLysLeuGlyAlaGlyThrLeuIleAlaAsnGlyGlnI 426
550 GGAACCAAGCTCGATCAGCTGGGCGGACGCTACAGTATTTTGATC 1441
551 yIleAsnGlnGlyAspIleSerIleGlyGlyGlyThrValValLeuAlaG 443
552 AGCAGGCGACGATAAAGGCAAAACACAGCTTTAGTGAATCGCTTG 1491
553 lnyAlaAlaSerAspGlySerLysGlnAlaPheAsnGlnValGlyIle 459
554 NTCAGGCGGAGGTACGCTGCAACTGAATGCGGATAATCAGTTCAACCC 1541
555 ThrSerGlyArgGlyThrAlaValLeuAlaAspSerGlnGlnIleLysPr 476
556 CGAACAACCTATTTCGGCTTCGCGGCGGACGTTTGGATTTAAACGGCG 1591
557 oGluAsnLeuTyrPheGlyPheArgGlyGlyArgLeuAspLeuAsnGlyA 493
558 ATTTCGCTTTCTCCCGCTATTCAAAATACCGATCAAGGGCGATGATT 1641
559 snAsnLeuAlaPheThrHisIleArgHisAlaAspGlyGlyAlaGlnIle 509
560 GNCNATCAATGCCCAACACATCCAGCTTACCATTACAGGGAATGA 1691
561 ValAsnHisAsnProAspGlnAlaAlaThrLeuThrLeuThrGlyAsnPr 526
562 A..... 1692

526 oValLeuSerProGluHisValGluTrpValGlnTrpGlyAsnArgProG 543
527 1692
528 lngAsnAlaAlaValTyrGluTyrIleAsnProHisArgAsnArgArg 559
529AGTATTACACACCGAGTGT..... 1713
530 ThrAspTyrPheIleLeuLysProGlyGlyAsnProArgGluPhePhePr 576
531AAGAATATC.....AATAGAC 1729
532 oLeuAsnMetLysAsnSerThrSerTrpGlnPheIleGlyAsnAsnArgG 593
533 TTAATTACAGCAAAAGAAATTGCC..... 1752
534 lngAlaAlaGluGlnValAlaGlnAlaGluAsnAlaArgProAspLeu 609
535 TACACGGTGTGGTGGCGAGAAAGATACGACC..... 1785
536 IleThrPheGlyGlyTyrLeuGlyGluAsnAlaGlnThrGlyLysAlaAl 626
537 1785
538 aProSerTyrSerLysThrAsnGluAlaAlaIleGluLysThrArgHisI 643
539AAACGACGGCGG 1800
540 leAlaAsnAlaAlaValTyrGlyArgProGluTyrArgTyrAsnGlyAla 659
541 CTCAACCTTGTTTACAGCCCGCCGAGAGACCGCCCGCTGCTTTC 1850
542 LeuAsnLeuHisTyrArgProLysArgThrAspSerThrLeuLeuLeuAs 676
543 CGCGGCAAAATTTAAACGCAACATCAGCAAAACAAACGCAAACTGT 1900
544 nglyGlyMetAsnLeuAsnGlyGluValLeuIleGlyGlyAsnMetI 693
545 TTTTACGGCGGACGACCGCGCTACATCATTTAGGAGCGCGG 1950
546 leValSerGlyArgProValProHisAlaTyrAspHisGlnAla..... 707
547 TGCTCAAAATGGAAGGTATCCCAAGGAGAAATCGTGTGGCAACGA 2000
548LysArgGluProValLeuGluAsnGl 716
549 CTGGATCACCACGCTTTAAAGCGGAAATTTCCATATTCAGGCG...G 2047
550 utrpThrAspGlySerPheLysAlaAlaArgPheThrLeuArgAsnHisA 733
551 GGCAGCGGTGATTTCCCGCAATGTTGCCAAAGTGAAGGCGATGNCAT 2097
552 laArgLeuThrAlaGlyArgAsnThrAlaHisLeuAspGlyAspIleThr 749
553 TTGACCAATCACGCCCAAGAGTGTTCGTCGACCGCCATCAAGGCA 2147
554 AlaTyrAspLeuSerGlyIleAspLeuGlyPheThrGlnGlyLysThrPr 766
555 TACAATCTGTACGCTTCGACGACGACGCTGACAAATTTGTGCGAAN 2197
556 oGlu...CysTyrArgSerTyrHisSerGlySerThrHisCys..... 779
557 AAACATTTACCGACGATAAAGTGATGTTGCTTCTTCACTTAAGCAGCAG 2247
558 779
559 AGCGGCANTGTNAGCTNNCCNATNACGNTNNTTAAACCTCNCNGGCN 2297
560 779
561 TGCNNCACTNAANGCAATCTTAGTGCAATGGCGATACAGTTATACAG 2347
562 779

853 CTTTAT.....TCGGCAGGGAACCGTTTCAGCTGATACGCCAA..893
||||| ::||| |||
293 ProTyrThrTyrTrpAlaGlyTyr-Glu.....LysLy 303

894 AGATTGG.....TTCTACGATCACATTTT 916
|:::||| |::| ||| |||
303 sSerTrpGlnGluTrpAsnIleTyrLysThrPheAlaAspGlyIleL 320

917 ACAGAGCGCATACACATACCGTGCTNTTTGAACCCGCAGTAACGGACAT 966
::: |||::: ||| ||||| ::::
320 ysAsnArgAspAsnAlalysProVal.....PropHeSerAsnLysGlu 334

967 TTTTCTTTTACATCAACAACAACAGGTACGGGTACGTTACAGAAACCAA 1016
::: ::|||::|||::|||::|||::|||::|||::|||::|||
335 TyrArgTrpThrAsnThrThrAsnHisGlnSerGluIleLysAsnThrAs 351

1017 CGAAAAGGTNTCCAATCCAAAGCTTAAAGTACACACATCCGACTGTTTG 1066
::::: ::: |::||| |||
351 phisThrIle.....ThrValThrLeu... 358

1067 ACGAATCTTTGAATGAAACTGATAAGAACAACAGTTTACGGCGACGGGGT 1116
:::~::~: |||::: |||
359ProSerAspProAspargLeuValAsnTyr..... 368

1117 GTTATACGTACCCTCCAAAGGTTTAAACAACCGGTGAACACCTTCCTTTTAT 1166
|||::: ::: |||::|||::|||::|||::|||::|||
369GlnLysGluLysAsnLysAsnThrGlyGlnAsnValIlePheGl 383

1167 CGATTACGCCAACGCCAA.....CTCATCTTATCAACAACATCAACC 1210
|||||::|||::|||::|||::|||::|||::|||::|||
383 uGlyAsnGlyAsnSerLysAsnThrLeuValLeuLysAsnIleAsnG 400

1211 AAGCGCGCGCGGTGTGATTTTGAAGTGATTTTACGGTCT...TCGCCCT 1257
|||||::|||::|||::|||::|||::|||::|||::|||
400 lnglYalaGlyLeuPhePheLysGlyAsnTyrGluValLysGlyThr 416

1258 GAAACAACAACCGTGGCAAGCGCGCGGTTCATATCAGTGAAGACAG 1307
:::~::~: |||::|||::|||::|||::|||::|||::|||
417 ThrAspAsnIleThrTrpValGlycylGlyIleAspValAlaGluGly 433

1308 TACCGTTACTTGGAAAGTAAACGGCGTGGCAACGACCGCTGTCCAAA 1357
:|||||::|||::|||::|||::|||::|||::|||::|||
433 sThrValThrTrpLysValHisAsnProGluLysAspHisLeuAlaLys 450

1358 TCGGCAAGGCAACGCTGCACGTTCAAGCCAAGGGGAAAAACCAAGGCTCG 1407
|||::|||::|||::|||::|||::|||::|||::|||
450 leglYlysGlyLysLeulleValGluGlyLysGlyAspAsnLysGlySer 466

1408 ATCAGCGTGGCGCAGGTACAGTCAATTTGGATCAGCAGCAGCAGATAA 1457
|||::|||::|||::|||::|||::|||::|||::|||
467 LeuLysValGlyAspGlyThrValValLeuLysGlnGlnThrThrThr.. 482

1458 AGGCAAAAACAACGCTTTAGTGAATCCGGCTTGTACGGCGACGGGGTA 1507
:::~::~: |||::|||::|||::|||::|||::|||::|||
483 ...GlyGlnHisAlaPheAlaSerValGlyIleValSerGlyArgSerT 498

1508 CGGTGCAACTGAATCCGATATATCAGTTCAACCCCGACAACCTATTTC 1557
|||::|||::|||::|||::|||::|||::|||::|||
498 hrValValLeuAsnAspAspyLysGlnValaspproAsnSerIleTyrPhe 514

1558 GCCTTCGCGCGGAGCTTTGGATTTAAACGGGCAATTCGGTTTGGTTCCA 1607
|||::|||::|||::|||::|||::|||::|||::|||
515 GlyPheArgGlyGlyArgLeuAspLeuAsnGlyAsnSerLeuthrPheAs 531

1608 CCGTATTCAAATACCGATGAAGGGCGATGATGNCNATCAATAATGCCA 1657
:::~::~: |||::|||::|||::|||::|||::|||::|||
531 phisIleArgAsnIleAspAspGlyAlaArgLeuValAsnHisAsnMetT 548

1658 CAACAACATCCACCGTTACCATTAACAGGGAATGAAGTATTACACACCG 1707
|||::|||::|||::|||::|||::|||::|||::|||
548 hrasAlaserAsnIlethrIlethrGlyGluSerLeuillethrAspPro 564

1708 AGT..... 1710

```

565  LysLysIleAsnProTyrTyrIleLysAlaArgGluGluAspAsnProTyr 581
1710 .....
581  rAlaPheArgTrpIleLysAspGlyGluInLeuTyrPheAsnLeuGluA 598
1710 .....
598  snTyrThrTyrAlaLeuArgLysGlyLysAlaAsnSerGlnLeu 614
1711 .....GSTRAGAA 1718
615  ProTyrAsnAspLysGluSerAsnGluAsnTrpLeuTyrMetGlyLysAs 631
1719 T.....ATCAATAGACTTAATACAGCAAG 1744
631  nAlaAspGluAlaLysArgAsnValMetLysHisIleAsnAsnGluArgM 648
1745  AAATGTCCTACAACGGTTGGTGGCAGAGAAGATACGACCAAAACGAAC 1794
648  etAsnGlyPheAsnGlyTyrPheGlyGluGluGlu..GlyLysAspAsn 663
1795  GGGCGCTCAACCTGTGTACCAGCCGCCGACGAAGACGCACCCNGCT 1844
664  GlyAsnLeuAsnValThrPheLysGlyLysThrGluGlnAsnArgPheLe 680
1845  GCTTTCGGCGGCAACAAATTTAAACGGCAACATCACGCAACAAACAGCA 1894
680  uLeuThrGlyGlyThrAsnLeuAsnGlyAsnLeuLysValGluGlnGlyT 697
1895  ACTGTGTTTTCAGCGCAGACGACACGCGACGCTACAATCATTTAGGA 1944
697  hrLeuPheLeuSerGlyArgProThrProHisAlaArgAspIleAlaGly 713
1945  ACGGGTGTCAAAATTCGAAGT.....ATCCCAACAGGAAATCGT 1988
714  IleSerSerThrLysLysAspSerHisPheAlaGluAsnGluValVa 730
1989  GTGGGCAACGACTCGATCNACGCGACGTTTAAAGCGGAAATTTCCATA 2038
730  ValGluAspAspTrpIleAsnArgAsnPheLysAlaThrThrMetAsnV 747
2039  TTCAGGCGGGCAGCGGTGATTCC...CGCAATGTGTCCAAAGTGGAA 2085
747  aThrGlyAsnAlaSerLeuTyrSerGlyArgAsnValGluSerIleThr 763
2086  GCGGATGNCATTTGAGCAATCAGCCCAACGACTTTTGGTGTCCGACC 2135
764  SerAsnIleThrAlaSerSerLysAlaGlnValHisIleGlyTyr.... 778
2136  GCATCAAAAGCCATACAANTCTGACAGTTCGGACTGACNGGTCTGCACAA 2185
779  LysLysGlyAspThrValCysValArgSerAspTyrThrGlyTyrValT 795
2186  ATTGTGTGCAANAANCAATTACCAGCAGTAAAGTGATTGCTTCATTGACT 2235
795  hrCys.....796
2236  AAGACGACGNTNAGCGGANTGTNAGCNTNCCNATNACGNTNNTNAAA 2285
796  .....796
2286  NCTCNCNGGCGNTGCNCCACTNNAANGGCAATCTTAGTCCAATGGCGAT. 2334
797  .....HisAsnGlyAspL 801
2335  ....ACAGGTTATACGTACGCCACCAACGCCACCCAAA...AACGGCAAC 2376
801  euSerGluLysAlaLeuAsnSerPheAsnAlaThrAsnValPheGlyAsn 817
2377  CTTAGCCTCGTGGCAATGCCCAAGCAACATTTTAAATCAAGCCACATATAA 2426

```



```

3465 ..... 3465
1344 aProAsnSerAlaGlnLysGluThrAlaGlnSerValAlaAsnProLysT 1361
3465 ..... 3465
1361 hrAlaGluGlnProValThrValSerThrGluAsnProValValGluAsn 1377
3466 .....CAACCTT.....CAACCCCA 3479
1378 ProGluAsnThrThrGlnProAlaValAsnSerGluAlaValGlnProGI 1394
3480 ..... 3480
1394 uThrAlaThrThrGlyThrValSerGlnProLysGluAlaSerThrAspG 1411
3480 ..... 3480
1411 luThrThrValAlaSerThrAspGluThrThrGlyThrSerAlaGluGlu 1427
3480 ..... 3480
1428 thrThrValAlaAspAsnSerGluAlaSerLysProLysArgSerAr 1444
3480 ..... 3480
1444 gArgAspValSerSerThrProHisAsnValGluProAlaValThrGlyG 1461
3481 .....CCGACGCGGACCTGATNAGC 3501
1461 lyGlyArgAspArgSerAlaValValProLeuArgAspLeuThrSer 1477
3502 CGTTATGCCAATAGCGGTTTGAGTCAATTTTCCGCC ..... 3537
1478 thrAsnThrAsnAlaValLeuSerAspAlaMetAlaLysAlaGlnPheVa 1494
3538 .ACGCTCAACAGCGTTTTCGCGGTACAGACCAATTTGGACGCGTG...T 3583
1494 lAlaLeuAsnValGlyLysAlaValSerGlnHisIleSerGlnLeuGluM 1511
3584 TTGCCGAAGACCGCGCGACGCGTTTGGACAAGCNGCATCCGGNACACC 3633
1511 eAsnAsnGluGlyGlnThrAsnValThrPValSerAsnThrSerMetAsn 1527
3634 AAACACTACCGTTCAGAGATTTCCGCGCCTACCGCCCAACAAACCGACCT 3683
1528 GluAsnThrSerSerAspGlnThrArgPheSerSerLysSerAlaGl 1544
3684 GCGCCAAATCGGTATGCAGAAACCTCGGACGCGG...CGGTGCGGCA 3730
1544 nThrGlnLeuGlyTrpAspGlnThrIleSerAsnAsnValGlnLeuGlyG 1561
3731 TCCTGTTCGCACACCGCGACCGGAAACANCTTCGACGACGCGATCGCG 3780
1561 lyIleLeuThrThrValArgAsnSerAsnSerPheAspLysAlaSerSer 1577
3781 AACTCGGACGCGTTCGCCACGCGCGCTTTTCGGGCAA...TACGGCAT 3827
1578 LysAsnThr...LeuAlaGlnAlaAsnPheThrLysThrAlaAs 1593
3828 CGCGAGTTCGACATCGGCATCAGCACGCGCGCGGTTTACGACGCGCA 3877
1593 pAsnHisTrpTrpLeuAlaValAspLeuGlyThrGlyAsnPheGlnSerA 1610
3878 NPTTCAGACGCGCATCGGAGCAAAATCCGCGCGCGGTGTCGATTCAC 3927
1610 snLeuGlnThrAsnHisAsnAlaLysPheAspArgHisThrAlaGlnIle 1626
3928 GGCATTACGACGACATACCGCGCGGTTCGCGGATTCGCGATCGAAC 3977
1627 GlyLeuThrAlaGlyLysAlaPheAsnLeuGlyAsnValAlaValLysPr 1643
3978 GTACATCGCGGCAACGCGCTATTTCGTCCAAAACGCGGATTACCGCTACG 4027

```

```

1643 oThrValGlyValArgTyrSerTyrLeuSerAsnAlaAspPheAlaLeuA 1660
4028 AAAACGTCATATATCGCCACCCCGGCTTCGCTTCAACCGNTACCGNGCG 4077
1660 spGlnAspArgIleLysValAsnProIleSerValLysThrAlaPheAla 1676
4078 GGCATTAAAGCAGATTATTTCATTCAACCGCGCACACATNTCCATCAC 4127
1677 GlnValAspLeuSerTyrThrTyrAsnLeuGlyGlu...PheAlaIleTh 1692
4128 NCCTTATTTNAGCCGTCCTATACCGATCGCGCTTCGGGCAAGTCCGAA 4177
1692 rProIleLeuSerAlaArgTyr...AspAlaAsnGlnGlyAsnGlyLysI 1708
4178 CAGCGCTCAATACCGCGGTATTGGCTCAGGATTTGGCAAAACCGCGAGT 4227
1708 leAsnValSerGlyTyrAlaPheAlaTyrAsnValGluAsnGlnGln 1724
4228 GCGAATGGGGCGTAACCGCGCAATCAAGGTTTCACGCTGTCNTCCA 4277
1725 TyrAsnAlaGlyLeuLysLeuLysTyrHisAsnValLysLeuSerLeuMe 1741
4278 CGTTCGCGCGCCAAAGNCCGCACTGGAAGCGCAACACAGCGCGGCA 4327
1741 tGlyGlyLeuThrLysAlaLysGlnAlaGlyLysGlnLysThrAlaGluV 1758
4328 TCAATTAGGTAC 4341
1758 allLysLeuSerPhe 1762

seq_name: sp_bacteriap:Q9JVB9
seq_documentation_block:
ID Q9JVB9 PRELIMINARY; PRT; 1773 AA.
AC Q9JVB9;
DT 01-OCT-2000 (TremBLrel. 15, Created)
DT 01-OCT-2000 (TremBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE IGAL PROTEASE (EC 3.4.21.7).
GN IGA OR NMA0905.
OS Neisseria meningitidis (serogroup A).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=65699;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Z2491 / SEROGROUP A / SEROTYPE 4A;
RX MEDLINE=20222556; PubMed=10761919;
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
RA Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holroyd S.,
RA Jagels K., Leather S., Moule S., Mungall K., Quail M.A.,
RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
RA Whitehead S., Spratt B.G., Barrell B.G.;
RT "Complete DNA sequence of a serogroup A strain of Neisseria
RT meningitidis Z2491."
RL Nature 404:502-506(2000).
DR EMBL; AL162754; CAB84182.1;
DR MEROPS; S06.001;
DR InterPro; IPR002195; Dihydroorotase.
DR InterPro; IPR000710; Iga_S6.
DR Pfam; PF02395; IGAL_1.
DR PRINTS; PRO0921; IGASERTASE.
DR PROSITE; PS00482; DIHYDROOROTASE_1; UNKNOWN_1.
KW Protease; Hydrolase; Complete proteome.
SQ SEQUENCE 1773 AA; 196350 MW; CAC19E7313D76CE1 CRC64;

```

```

alignment_scores:
Quality: 1410.00      Length: 1849
Ratio: 1.495          Gaps: 60
Percent Similarity: 51.001 Percent Identity: 26.501

```



```
544 erValTyrGluIleArgAsnAspTyrAspAspAspTyrTyrGlyTyr 560
1705 .....CCGAGTGGTAAGAATAATATAGACTTAA 1733
561 TyrSerTyrArgLysProIleProGlnGlyLysAspLeuTyrTyrLysAs 577
1734 TTAC..... 1737
577 nTyrArgTyrTyrAlaLeuLysSerGlyGlySerValAsnAlaProMetp 594
1737 ..... 1737
594 roGluAsnGlyGlnThrGluAsnAsnAspTrpIleLeuMetGlySerThr 610
1738 .....AGCAAGAAATTCCTACAAC..... 1758
611 GlnGluGluAlaLysLysAsnAlaMetAsnHisLysAsnAsnGlnArgII 627
1759 .....GGTTGGTTGGCGAGAAGATACGACCAAAACGCAACG 1795
627 eSerGlyPheSerGlyPheGlyGluGluAsnGlyLysGlyHisAsnG 644
1796 GCGGCTCAACCTTGTTCACAGCCCGCCGACAGACCCACCCGCGCTG 1845
644 lAlaLeuAsnLeuAsnPheAsnGlyLysSerAlaGlnAsnArgPheLeu 660
1846 CTTTCCGCGGCAACAAATTTAAACGGCAACATCAGCGCAACAAACGCA 1895
661 LeuThrGlyGlyThrAsnLeuAsnGlyLysIleSerValThrGlnGlyAs 677
1896 ACTGTTTTCAGCGGCGAGACCGACCGCGCTCAATCAATATTAGGAA 1945
677 nValLeuLeuSerGlyArgProThrProHisAlaArgAspPheValAsnL 694
1946 CCGGCTGTCAAAATCGAAGT.....ATCCACAAAGAGAAATCCTG 1989
694 ysSerSerAlaGlnLysAspAlaHisPheSerLysAsnAsnGluValVal 710
1990 TGGGACACGACTGGATCNACCGCAGCTTTAAAGCGGAAATTTCCATAT 2039
711 PheGluAspAspTrpIleAsnArgThrPheLysAlaThrGluIleThrVa 727
2040 T...CAGGCGGCGAGCGGTGATTCCCGCAATGTTGCCAAAGTGAAG 2086
727 lAsnGlnSerAlaSerPheSerSerGlyArgAsnValSerAsnIleThrA 744
2087 GCGATTGNCATTTTCAGCAATCAGCCCAACGAGTGTGTCGCAACG 2136
744 lAsnIleThrAlaThrAspAsnAlaLysValAsnLeuGlyTyr..... 758
2137 CATCAAGCCCATACAATCTGTACACGTTTCGACTGGACNGTCTGACAAA 2186
759 LysAsnGlyAspGluValCysValArgSerAspTyrThrGlyTyrValTh 775
2187 TTGTGCTGAANAANCATTACCGAGATAAAGTGATTCTTCAATGACTA 2236
775 rCysAsnThrGlyAsnLeuSer...AspLysAlaLeuAsnSerPheGlyA 791
2237 AGACNGACNTNAGCGGCANTNAGCTNAGCTNCCNATNAGTNTNTNAAAN 2286
791 laThrLysIleAsnGlyAsnValAsnLeuAsnGlnAsnAlaLeuVal 807
2287 CTCNCGGCGCTGNNCAGTNAAGCAATCTTAGTGCAATGGCGATAC 2336
808 Leu...GlyLysAlaAlaLeuTrpGlyGlnIleGlnGlyGlnGlyAsnSe 823
2337 ACGTTATACACTAGCCACACGCGC.....ACCC 2365
823 rArgValSerLeuAsnGlnHisSerLysTrpHisLeuThrGlyAspSerG 840
2366 AAAACGGCAACCTTAGCCTC.....GTG 2388
|| |||||
```

```
840 lNValHisAsnLeuSerLeuAlaAspSerHisIleHisLeuAsnAla 856
2389 GCAATGTCACCAAGCAACATTTAATCAAGCCACATTAACGCGCAACNCATC 2438
857 SerAspAlaGlnSerAlaAsnGlnTyrHisThrLeuLysIleAsn...Hi 872
2439 GGNTCGGGCAATGCTTCATTAATCTAAGCAACACCGCGCACAAAC. 2487
872 sLeuSerGlyAsnGlyHisPheHisTyrLeuThrHisLeuAlaGluAsnL 889
2487 ..... 2487
889 euGlyAspLysValLeuValLysGluSerAlaSerGlyHisTyrGlnLeu 905
2488 .....GGCAGTCTGACGCTTTC 2504
906 HisValGlnAspLysThrGlyGluProAsnGlnGlyLysLeuAsnLeuPh 922
2505 CGACACAGCT.....AAGGCAACGTAAGCCATTCGCGCACTCA 2542
922 eAspAlaSerSerValArgAspArgSerHisLeuSerValSerLeuAla 939
2543 ACGGCAATGCTCTCCCTAGCGGATAAGGAGTATTCATT..... 2581
939 snHisValAspLeuGlyAlaLeuArgTyrThrIleLysThrGluAs 955
2582 .....TTGA 2585
955 nGlyIleThrArgLeuTyrAsnProTyrAlaGluAsnArgArgValL 972
2586 AAACAGCCGCTTTACGGACAACCT..... 2609
972 ysProAlaProSerProAlaThrAsnThrAlaSerGlnAlaGlnThrAs 988
2609 ..... 2609
988 pSerAlaGlnIleAlaLysProGlnAsnIleValValAlaProProSerP 1005
2610 .....CAGCGGCGCAAGGACAGCAT 2633
1005 roGlnAlaAsnGlnAlaGluGluAlaLysArgGlnGlnAlaLysAlaGlu 1021
2634 ACATTTAAAGACAGCGAATGGAC..... 2657
1022 GlnValLysArgGlnGlnAlaGluAlaGluArgLysSerAlaGluLeuAl 1038
2658 .....GCTGCGCTCAGCGACGGAATTAGCAATTTTAAACCTTGAC 2697
1038 aLysGlnLysAlaGluAlaGluArgGluAlaArgGluLeuAla.....T 1053
2698 AACGCCACCATTAACCTCAATTCGCGCTATCCGACGATGTCGAGCGC 2747
1053 hrArgGlnLysAlaGluGlnGluArgSerSerAlaGluLeuAlaArgArg 1069
2748 GCNAACCGCGCAGNGTGTACAGACCGCGCGCGCTTCGCGCGCTTCCC 2797
1070 HisGluLysGlu.....ArgGluAlaAlaGluLeuSerAlaLysGlnLy 1084
2798 TATT.....ATCCGTTTACCCGCCA 2817
1084 sValGluAlaGluArgGluAlaGlnAlaLeuAlaValArgArgLysAlaG 1101
2818 ACTTCGGTAGAATCCGTTTCAACACGCTGACGGTAAACGGCAATTTGAA 2867
1101 luAlaGluGluAlaLysArgGlnAlaAlaGluLeuAlaArgArgHisGlu 1117
2868 CNGTCAAGCAACATTCGCGCTTTATGTCGGAACCTCTTCGCTA..... 2909
1118 LysGluArgGluAlaAlaGluLeuSerAlaLysGlnArgValGlyGluG 1134
2910 .CCGAGCGCACAAATTTGAAGCTGGCGGAAGTTCCGAAGGNACTTACACC 2958
1134 uGluArgArgGlnThrAlaGlnSerGlnProGlnArgLysArgArgA 1151
|| |||||
```


2959 TTGGCGGTCAACATAC.....CGGCAACGAAACCCGT 2990
1151 laAlaProGlnAspTyrMetAlaAlaSerGlnAspArgProLysArgArg 1167
2991 AAGCCTCATCAATTGACGGTAGTGGAGGAA..... 3023
1168 GlyHisArgSerValGlnGlnAsnValGluLeuAlaGlnAla 1184
3024AGACAAACAA..... 3032
1184 aGluLeuAlaArgGlnGlnGluGluArgLysAlaAlaGluLeuAla 1201
3033ACCGGTGTC 3041
1201 laLysGlnArgAlaGluAlaGluArgGluAlaGlnAlaLeuAlaAlaArg 1217
3042 CGAAACCTTAATTACCTCGCAAAACGACACGTCGATGCCGGCGCGT 3091
1218 ArgLysAlaGluAlaGluGluAlaLysArg, GlnAlaAlaGluLeuAlaH 1234
3092 GGCCTTACCACTCATCCGCAACACGGCGAGTTCCCGCTGCATAAT... 3138
1234 isArgGlnGluAlaGluArgLysAlaAlaGluLeuSerAlaAsnGlnLys 1250
3139 ...CCGGTCAAGAACAGAGCTTCGCACAAACTCGCAAGCAGAGAAGC 3185
1251 AlaAlaAlaGluAlaGlnAlaLeuAlaAlaArgGlnGlnLysAlaLeuAl 1267
3186 CAAAAACAGCGGAGAAAGACACACGCGCAAGCCTT..... 3222
1267 aArgGlnGlnGluAlaGluAlaArgLysAlaAlaGluLeuAlaValLysGlnL 1284
3223GACCGCTGATTGCGGCGCGCGCGATGCC 3252
1284 ysAlaGluThrGluArgLysThraAlaGluLeuAlaLysGlnArgAlaAla 1300
3253 GCCGAA.....AAGACAGAAAGCTTCGCGAACCGGC 3284
1301 AlaGluAlaLysArgGlnGlnGluAlaArgGlnThraAlaGluLeuAl 1317
3285 CCGCGCGAGCGGGGAAATGTCGCATTATGACGCGGAGGAA... 3330
1317 aArgArgGlnGluAlaGluArg.....GlnAlaAlaGluLeuS 1330
3331 ..GAGAAAAACGGTGCAGCGGATAAACACACCGCCTTGGCGAARACAG 3378
1330 erAlaLysGlnLysAlaGluThrAspArgGluAlaAlaGluSerAlaLys 1346
3379 CGGAAGCGGAAACCCCGCGGNTACCACCGCCTTC..... 3414
1347 ArgLysAlaGluGluGluHisArgGlnAlaAlaGlnSerGlnProG 1363
3414 3414
1363 nArgArgLysArgArgAlaAlaProGlnAspTyrMetAlaAlaSerGlnA 1380
3415CCCCGCGCGCGCGCGCGGATTCGCGCAACCG..... 3453
1380 snArgProLysArgArgGlyArgArgSerThrLeuProAlaProSer 1396
3453 3453
1397 ProSerPheAspSerSerAlaTyrAlaAlaProArgAlaLeuHisAsnPr 1413
3453 3453
1413 oAspTrpTyrGluAsnAspTyrGluGluIleProLeuAspAlaLeuGluA 1430
3453 3453
1430 spGluAsnValSerGluSerValAspThrSerAspLysGlnProGlnAsp 1446
3453 3453
1447 AsnThrGluLeuHisGluLysValGluThrValSerLeuGlnProArgAl 1463
3454CAGCCC.....CACCAGCACTCAACCCCAAC 3481
1463 aAlaGlnProArgAlaGlnAlaAlaThrGlnProGlnAlaGlnAlaAla 1480
3482 CCAGCGCGACCTGATNAGCCGTTATGCAATAGCGTTTGGAGTAATTT 3531
1480 laGlnAlaAspAlaValSerThrAsnThrAsnSerAlaLeuSerAspAla 1496
3532 TCCGACACGCTCAACAGCGTTTTCGCC...GTACAGGACGATTCGACCG 3578
1497 MetAlaSerThrGlnSerIleLeuAspThrGlyAlaSerLeuThrAr 1513
3579 COTGTTTGGCGAAGACCGCGC.....AACGCGTTTGA 3613
1513 gHisIleAlaGlnLysSerArgAlaAspAlaGluLysAsnSerValTpm 1530
3614 CAAGNGCATCCGACCAACACACTACCGTTCCGCAAGATTCGCGCGC 3663
1530 etSerAsnIleGlyTyrGlyArgAspTyrAlaSerAlaGlnTyrArgArg 1546
3664 TACCGCCAAACACGACCTGCGCAAAATCGGTATCGCAAAACCTCGG 3713
1547 PheSerSerLysArgThrGlnThrGlnIleGlyIleAspArgSerLeuSe 1563
3714 CAGCGGG...CGGTGCGCATCTGTTTCGCACAAACCGGACCGGAAACA 3760
1563 rGluAsnMetGlnIleGlyGlyValLeuThrTyrSerAspSerGlnHis 1580
3761 NCTTCGACGACGCGCATCGCAACTCGGCACGCTTCCGCGCGCGCTT 3810
1580 hrPheAspGlnAlaSerGlyLysAsnThr...PheValGlnAlaAsnLeu 1595
3811 TTCGGCAATACGGCATCGCGAGTTCGACATCGGCATCGACGCGCGC 3860
1596 TyrGlyLysTyrLeu.....AsnAspAl 1604
3861 GGTGTTTACGCGGANTCTNTACAGCGCATCGGAGGCAAAATCCGCG 3910
1604 aTrpTyrMetAlaGlyAspIleGly.....AlaGlySerLeuArgS 1618
3911 GCGCGCTGCTGCATTACGCGCATTCAGGCACGATACGCGCGCGTTC... 3957
1618 erArg.....LeuGlnThrGlnGlnLysAlaAsnPheAsn 1629
3958GG 3959
1630 ArgThrSerIleGlnThrGlyLeuThrLeuGlyAsnThrLeuLysIleAs 1646
3960 CGGATTGCGCATCGAACCGTACATCGCGCAACCGCGCTATTTCGTCCAA 4009
1646 nGlnPheGluIleValProSerAlaGlyIleArgTyrSerArgLeuSerS 1663
4010 AACCGGATTACCGCTACGAAACGTCATATCGCCACCCCGCTTCCG 4059
1663 erAlaAspTyrLysLeuGlyAsnAspSerValLysValSerSerMetSer 1679
4060 TTCACCGGNTACCGCGCGCATTAAGCAGATTTATTCATCAACCGCGC 4109
1680 valLysThrLeuThrAlaGlyLeuAspPheAlaTyrArgPheLys...Va 1695
4110 GCAACACATNTCCATCACNCCTTATTNNAGCCTGCTCTAT...ACCGATG 4156
1695 lGlyAsnLeuThrValLysProLeuLeuSerAlaAlaTyrPheAlaAsnT 1712
4157 CGCCTTCGGGCAAGTCCGACACCGCGTCAATACCGCNGTATTGGCTCAG 4206
1712 yrGlyLysGlyGlyValAsnValGlyGlyAsnSerPheValTyrLysAla 1728
4207 GATTTTCGCAAAACCGCGCAGTGGGAATGGCGTAAACCGCGCAATCAA 4256

```

1729 AspAsnGlnGlnTyrSerAla.....GlyAlaAlaLeuLeuTyrAr 1743
4257 AGGTTTCACGCTGTCNPTCCACGCTGCGCGCAAAAGGCCGCAACTGG 4306
1743 GAsnValThrLeuAsnValAsnGlySerIleThrLysGlyLysGlnLeuG 1760
4307 AAGCGCAACACAGCGCGGCATCAATATAGCTACCGCTGG 4347
1760 LuLysGlnLysSerGlyGlnIleLysIleGlnIleArgPhe 1773

```

seq_name: sp_bacteriap:Q9K0B4

seq_documentation_block:

```

ID Q9K0B4 PRELIMINARY; PRT; 1815 AA.
AC Q9K0B4;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
DE IGA-SPECIFIC SERINE ENDOPEPTIDASE.
GN NMB0700.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=MC58 / SEROGROUP B;
RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA Nelson W.C., Gwinn S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Haft D.H., Salzberg S.L., Deboy R., Peterson J.D., Hickey E.K.,
RA Mason T., Clecko A., Parksey D.S., Blair E., Ciftone H., Clark E.B.,
RA Cotton M.D., Ufferback T.R., Khouri H., Qin H., Vamathevan J.,
RA Gill J., Scarlato V., Maignani V., Pizzo M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
MC58."
RL Science 287:1809-1815(2000).
DR EMBL; AE002424; AAF41117.1; -.
DR MEROPS; S06.001; -.
DR TIGR; NMB0700; -.
DR InterPro; IPR002195; Dihydroorotase.
DR InterPro; IPR000710; IGA_S6.
DR Pfam; PF02395; IGA1; 1.
DR PRINTS; PR00921; IGASERPTASE.
DR PROSITE; PS00482; DIHYDROOROTASE_1; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 1815 AA; 201077 MW; 2259D4D71762C57F CRC64;

```

alignment_scores:

```

Quality: 1404.00 Length: 1878
Ratio: 1.479 Gaps: 60
Percent Similarity: 50.532 Percent Identity: 26.198

```

alignment_block:

US-09-303-518d-651 x Q9K0B4 ..

Align seg 1/1 to: Q9K0B4 from: 1 to: 1815

```

49 AAAACCGCGCGCATCCGCTTCTCGCTGCTTACTTACCCATATGCCGTGC 98
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2 LysThrLysArgPheLysIleAsnAlaIleSerLeuSerIlePheLeuAl 18
:|||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||
99 GTTCGGCATCTTCTCCCAAGCTTGGCGGGGACACACTTATTTCGGCATCA 148
:|||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||
18 aTyrAlaLeuThrProTyrSerGluAlaAlaLeuValArgAspVala 35
:|||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||
149 ACTACCAATACTATCGCGACTTTGCCGAAATAAAGCAAGTTTGCAGTC 198
:|||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||
35 sPtyrGlnIlePheArgAspPheAlaGluAsnLysGlyLysPhePheVal 51
:|||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||

```

```

199 GGGCGCAAGATATATTGAGTNTACACAAAAAGGGGAGTTGGTCGGCAA 248
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
52 GlyAlaThrAspLeuSerValLysAsnLysGlnGlyAsnIleGlyAs 68
249 ATCAATGACAAAGCCCGCATGATGATTTTCTGTGTGTGTGTGTGTGTAAC 297
:|||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||
68 nAlaLeuSerAsnValProMetIleAspPheSerValAlaAspValAsnA 85
298 ..GGCGTGGCGCATGTTGGTGGCGATATATATTGAGCGCTGGCACAT 345
:|||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||
85 rgArgThrLeuThrValIleAspProGlnTyrAlaValSerValLysHis 101
346 AAGCGCGCTATAACAAACGTTGATTGTCGCGGAAGGAAGNAATCCCGA 395
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
102 ValLysGlyAspGluIleSerTyrTyrGlyHisHisAsnGlyHisLeuAs 118
396 T.....CAGCACCGCTTTTCTTACCAATTTGTGAAAGCAAAATATTATA 439
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
118 pValSerAsnAspGluAsnGluTyrArgSerValAlaGlnAsnAspTyrG 135
440 AGCTGCACAATTCA.....CACCCCTTACAACGGC.....GATTAN 474
:|||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||
135 LuProAsnLysAsnTrpHisHisGlyAsnGlnGlyArgLeuGluAspTyr 151
475 CATATGCCCGCTTGCATAAATTTGTACAGATGCAGAACCTGTGCAAAAT 524
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
152 AsnMetAlaArgLeuAsnLysPheValThrGluValAlaProIleAlaPr 168
525 GACGAGTGCATCAGGGGG.....AATACCTATTCCGATTAAGAAATAATC 571
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
168 oThrSerAlaGlyGlyValGluThrTyrLysAspLysAsnArgPheS 185
572 CGGAGCGTCTCGCATCGCTCAGGACACACTATTGGCGTTATGATGAT 621
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
185 erGluPheValargValGlyAlaGlyThrGlnPheGluTyrAsnSerArg 201
622 GACAAACAGCGGATTTATCTACTCCGCGCATGTTAATTTGGCGGAA 671
:|||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||
202 TyrAsnMetThrGluLeuSerArgAlaTyrArgTyrAlaIleAlaGlyTh 218
672 TACACATATGCAG..... 684
:|||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||
218 rProTyrGlnAspValAsnValThrSerAsnLeuAsnGlnGlyLeuI 235
685 ..GGTTGGGAAATAATGGCGTANTTAGTTTGAGCGCGCATGTGCGCAT 732
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
235 leGlyPheGlyAspAsnSerLysHisHisSerProGluLysLeuLysGlu 251
733 GCCAAGCACTATGCCCTATGCGGATTCAGGTGCGGCGGCGGACAGCGG 782
:|||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||
252 ValLeuSerGlnAsnAlaLeuThrAsnTyrAlaValLeuLysAspSerGl 268
783 TTCGCCAATGTTTATTATGACAAACAAACAAATAAATGGCTGCTCAACG 832
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
268 ySerProLeuPheAlaTyrAspLysGlnGlyLysArgTrpValPheLeuG 285
833 GAGTTTTCAAA.....ACCGGCTACCCTTATTCGCGCAGGAGGAAACGGT 876
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
285 lYalaTyrAspTyrTrpAlaGlyTyr.....GlnLysAsnSer 297
877 TTCCAG.....CTGATACGCAAGATTTGGTTCTACGATGACATTT...TA 917
:|||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||
298 TrpGlnGluTrpAsnIleTyrLysGluPheAlaAspGluIleLysGl 314
918 CAGAGCGGATACACATACGCTCTNTTTTGAACCGCGCAGTAACGGA... 963
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
314 nArgAspAsnAlaGlyThrIle.....LysGlyAsnGlyGluH 327
964 ..CATTTTTCCTTTACATCCACAAACAAACGTTACGGGTACGGTACACAGAA 1011
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
327 lshIstPlystThrThrGlyThrAsnSerHisIleGlySerThr..... 341
1012 ACCAAGAAAGGTTTCCAAAGCTTAAAGTACACAGACAGTCCGACT 1061

```

```
342 .....:::|||||
1062 GTTTGACGAATCTTTGAATGAACACTGATAAAGAACCAAGTTTACCGCGCAG 1111
345 ualaasn.....AsnGluArgAsp.....351
1112 GGGGTGTTAATCAGTACCGTCCCAAGGTTAAACCAACGGTGAACCTTTCT 1161
352 .....:::|||||AlaAsnAsnGlyGlnAsnValThr 359
1162 TTTATCATGATACGGCAGCAACGAACACTCATCTTATCAACCAACATCAACCA 1211
360 Phe.....GluAsnAsnGlyThrLeuValLeuAspGlnAsnIleAsnGI 374
1212 AGCGCGGGGGTGTGATTTGAGGTGATTTACGGTCTCGCTCGAA 1261
374 nglyAlaGlyLeuPhePheLysGlyAspTyrThrValLysGlyIleA 391
1262 ACAACGAA...ACGTGCAAGGGCGCGCTTCATATCATGTCAGTGAAGACAGT 1308
391 snAsnAspIleThrTriPheGlyAlaGlyIleAspValAlaAspGlyLys 407
1309 ACCGTTACTTGGAAAGTAAACGGCGTGGCAACGACCGCTGTCACAAAT 1358
408 LysValValTrpGlnValLysAsnProAsnGlyAspArgLeuAlaLysII 424
1359 CGGCAAAAGGCACGCTGCACGTTCAAGCCAAAGGGGAAACCAAGCTCGA 1408
424 eglyLysGlyThrLeuGluIleAsnGlyThrGlyValAsnGlnGlyLnl 441
1409 TCAGCGTGGCGCGCTGACAGTTCATTTGGATCAGCAGGCGAGACGATAAA 1458
441 euLysValGlyAspGlyThrValIleLeuAsnGlnGlnAlaAspAlaAsp 457
1459 GGCAAAAAACAAGCCTTTAGTGAATCGGCTGNTCAGCGGCGAGGGGTAC 1508
458 LysLysValGlnAlaPheSerGlnValGlyIleValSerGlyArgGlyTh 474
1509 GGTGCACTGAATCCGATTAATCAGTTCACCCCGCAAACTCATTTTCG 1558
474 rLeuValLeuAsnSerSerGlnIleAsnProAspAsnLeuTyrPheG 491
1559 GCTTTCGCGCGGACGTTTGAATTAACGGCATTTCGTTTCGTTCCAC 1608
491 lyPheArgGlyGlyArgLeuAspAlaAsnGlyAsnAspLeuThrPheGlu 507
1609 CGTATTCAAATACCGATGAAGGGCGCATGATTGNCNATCATATAATGCCAC 1658
508 HisIleArgAsnValAspGluGlyAlaArgIleValAsnHisAsnThrGI 524
1659 AACACATCCACCGTTACCATTAACAGGAATGAAGTATTACAA... 1704
524 yHisAlaSerThrIleThrLeuThrGlyLysSerLeuIleThrAspProL 541
1704 .....:::|||||
541 ysThrIleSerIleHisTyrIleGlnAsnAsnAspAspAspAlaGly 557
1705 .....:::|||||CCGAGTGGTAAGAAATATC... 1722
558 TyrTyrTyrTyrArgProArgLysProIleProGlnGlyLysAspLeuTy 574
1722 .....:::|||||
574 rPheLysAsnTyrArgTyrTyrAlaLeuLysSerGlySerValAsnA 591
1722 .....:::|||||
591 laPrometProGluAsnGlyGlnThrGluAsnAsnAspTrpIleLeuMet 607
1723 .....:::|||||AATAGACTTAATTA 1736
:::|||||
```

```
608 GlySerThrGlnGluAlaLysLysAsnAlaMetAsnHisLysAsnAs 624
1737 CAGCAAGAATAATGCTTACACAGCTTGGTTTGGCGAGAAAGATACGACCA 1786
624 ngInArgIleSerGlyPheSerGlyPheGlyGluGluAsnGlyLysG 641
1787 AAACGAACGGCGGTCAACCTTTACAGCGCCCGCGCAGAACAGCGC 1836
641 lyHisAsnGlyAlaLeuAsnLeuAsnPheAsnGlyLysSerAlaGlnAsn 657
1837 ACCCGCTGCTTTCCGGGGAACAAATTTAAACGGCAACATCACGCAAA 1886
658 ArgPheLeuLeuThrGlyGlyThrAsnLeuAsnGlyLysIleSerValTh 674
1887 AAACGGCAAACTGTTTTCAGCGGAGACCGGACCGCAGCGCTACATC 1936
674 rgInGlyAsnValLeuLeuSerGlyArgProThrProHisAlaArgAsp 691
1937 ATTTAGGAAGCGGTGGTCAAAAATGGAAGT.....ATCCACAAGGA 1980
691 heValAsnLysSerSerAlaArgLysAspAlaHisPheSerLysAsnAsn 707
1981 GAAATCGTGTGGCAACGACTGATCACCACGCGCTTTTAAAGCGGAA 2030
708 GluValValPheGluAspAspTrpIleAsnArgThrPheLysAlaThrGI 724
2031 TTTCCATATT...CAGGGGGCGGCGGTGATTTCGCCGAATGTTGCCA 2077
724 utIleAlaValAsnGlnSerAlaSerPheSerSerGlyArgAsnValSerA 741
2078 AAGTGAAGCGGCTGTCATTTGACCAATCAGCGCCCAAGCAGTTCGTT 2127
741 spIleThrAlaAsnIleThrAlaThrAspAsnAlaLysValAsnLeuGly 757
2128 GTCGACCGCATCAAGGCATACAACTCTGTACACGTTGCGACTGACNGG 2177
758 Tyr.....LysAsnGlyAspGluValCysValArgSerAspTyrThrGI 772
2178 TCTGCAAAATGCTGCGAANAANCATTACCGACGATAAAGTGATGCTT 2227
772 yTyrValThrCysAsnThrGlyAsnLeuSer...AspLysAlaLeuAsn 788
2228 CATTCACATAAGCAGCANTAGCGGCANTGTNAGNCTNCCNATNACONT 2277
788 exPheGlyAlaThrGlnIleAsnGlyAsnValAsnLeuAsnGlnAsnAla 804
2278 NNTTAAANCTCNCNGCGCTGNNCACTNAANGCAATCTTAGTGCAA 2327
805 AlaLeuValLeu...GlyLysAlaAlaLeuTrpGlyGlnIleGlnGlyGI 820
2328 TGGCATACAGTTATACAGTCAGCCACACGCC..... 2361
820 nglyAsnSerArgValSerLeuAsnGlnHisSerLysTrpHisLeuThrG 837
2362 .....ACCCAAAACGGCAACCTTAGCCTC..... 2385
837 lyAspSerGlnValHisAsnLeuSerLeuAlaAspSerHisIleHisLeu 853
2386 .....GTGGGCAATGCCCAAGCAACATTTAATCAAGCCACATTAACGG 2429
854 AsnAsnAlaSerAspAlaGlnSerAlaAsnLysTyrHisThrLeuLysII 870
2430 CAACNCATCGGNTTCGGCAATGCTTCATTTAATCTAAGCAACACGCCG 2479
870 easn...HisLeuSerGlyAsnGlyHisPheHisTyrLeuThrHisLeuA 886
2480 CACAAAAC..... 2487
886 laLysAsnLeuGlyAspLysValLeuValLysGluSerAlaSerGlyHis 902
2488 .....:::|||||GGCAGTCT 2495
903 TyrGlnLeuHisValGlnAspLysThrGlyGluProAsnGlnGlyLe 919
```

2496 GACGCTTCCGACACGCTAAGCAAACGTAAGCCATT 2533
|::||| ||| :::::|||||
919 uAsnLeuPheAspAlaSerValGlnAspArgSerArgLeuSerValS 936
2534 CCGCACTCAACGGCAATCTCCTACCGCATAGGAGCAGTATCCATT... 2581
|::|||:::|||||:::|||||:::|||||:::|||||:::|||||
936 eLeuAlaAsnAsnHisValAspLeuGlyAlaLeuArgTyrThrIle 952
2581 2581
952 sThrGluAsnGlyIleThrArgLeuTyrAsnProTyrAlaGluAsnArg 969
2592TTGAAACAGCCGCTTTACCGACAACCT 2609
:::||||| ||| :::::|||||
969 rgArgValLysProAlaProSerProAlaThrAsnThrAlaSerGlnAl 985
2609 2609
985 aGlnLysAlaThrGlnThrAspGlyAlaGlnIleAlaLysProGlnAsnI 1002
2610CAG 2612
:::
1002 leValValAlaProSerProGlnAlaAsnGlnAlaGluGluAlaLys 1018
2613 CGGCAGCAGGACAGCATTTACACTTAAAGACAGCAGCATGGAC 2657
|||||::: ||| :::::|||||::: |||
1019 ArgGlnGlnAlaLysAlaGluGlnValLysArgGlnGlnAlaGluAlaG 1035
2658GCTGCGCTCAGGCACGGAA 2676
||| :::::|||||
1035 uArgLysSerAlaGluLeuAlaLysGlnLysAlaGluAlaGluArgGlu 1052
2677 TTAGCAATTTAAACCTTGACACGCCACCATTTACACTCAATTCGCGCTA 2726
|||::: |||::: |||::: |||::: |||::: |||::: |||
1052 laArgGluLeuAla.....ThrArgGlnLysAlaGluGluArgSer 1066
2727 TCGCCACATGCTCGACGGCGCAACCGGACGNGTGTACACACGCGCG 2776
|||::: |||::: |||::: |||::: |||::: |||::: |||
1067 SerAlaGluLeuAlaArgArgHisGluLysGlu.....ArgGluAlaAl 1081
2777 GCGCGCTCGCGCTTCCCTATT 2801
|||
1081 aGluLeuSerAlaLysGlnLysValGluAlaGluArgGluAlaGlnAla 1098
2802ATCCGTTACACCGCACTTCGGTAGAATCCGTTTCAACAGCCT 2846
:::||||| |||::: |||::: |||::: |||::: |||
1098 euAlaValArgArgLysAlaGluAlaGluGluAlaLysArgGlnAlaAla 1114
2847 GACGGTAACGGCAATTTGACNGTCAAGGAACATTCGGCTTTATGTCGG 2896
|||::: |||::: |||::: |||::: |||::: |||::: |||
1115 GluLeuAlaArgArgHisGluLysGluArgGluAlaAlaGluLeuSerAl 1131
2897 AACTCTTCGGCTACGAGCGCAAAATTAAGCTCGCGGAA 2937
:::|||||::: |||::: |||::: |||::: |||::: |||
1131 aLysGlnArgValGlyGluGluArgArgGlnThrAlaGlnSerGlnp 1148
2938 AGTCCGAGGNACTTACACCTTGGCGGTCAACAATAC 2975
|||::: |||::: |||::: |||::: |||::: |||
1148 roGlnArgArgLysArgAlaAlaProGlnAspTyrMetAlaAlaSer 1164
2976CGGCAACGACCCGTAAGCCTCGATCAATTGACGGTAGTGAAG 3019
|||::: |||::: |||::: |||::: |||::: |||
1165 GlnAspArgProLysArgArgGlyHisArgSerValGlnGlnAsnAsnVa 1181
3020 GGAAAGACAACA 3032
|||
1181 lGluIleAlaGlnAlaGlnAlaGluLeuAlaArgGlnGlnGluAla 1198
3032 3032
1198 rgLysAlaAlaGluLeuAlaLysGlnArgAlaGluAlaGluArgGlu 1214
3033ACCGTGTCCGAAACCTTAATTTACCCCTGCAAAACG 3070
:::||||| |||::: |||::: |||::: |||::: |||
1215 AlaGlnAlaLeuAlaAlaArgLysAlaGluAlaGluGluAlaAlaSer 1231
3071 AACACGTCGATCCGCGCGTGTACCACTCATCCGCAAAACGCGC 3120
|:::||||| |||::: |||::: |||::: |||::: |||
1231 g.GlnAlaAlaGluLeuAlaHisArgGlnGluAlaGluArgLysAlaAla 1247
3121 GAGTTCGCGCTGCATAATCCGTCAAAAGAACAGAGCTTCCGA 3164
|||::: |||::: |||::: |||::: |||::: |||
1248 GluLeuSerAlaAsnGlnLysAlaAlaGluAlaGlnAlaLeuAlaAl 1264
3165 CAAACTCGCGCAGAGCAAGCAAAACAGCGCGGAAAGAACACGCGC 3214
:::||||| |||::: |||::: |||::: |||::: |||
1264 aArgGlnGlnLysAlaLeuAlaArgGlnGlnGluAlaArgLysAlaAla 1281
3215 AAAGCCTTGACGCGCTG 3231
:::|||||
1281 laGluLeuAlaValLysGlnLysAlaGluThrGluArgLysThrAlaGlu 1297
3232 ATTGCGCGCGCGCGATGCCCGCAAAAGAC 3263
:::||||| |||::: |||::: |||::: |||::: |||
1298 LeuAlaLysGlnArgAlaAlaGluAlaLysArgGlnGlnGluAl 1314
3264 AGAAACGCTGCGCAACCGCGCGCGCGGCAAGGAAATGTCGCA 3313
|||::: |||::: |||::: |||::: |||::: |||
1314 aArgGlnThrAlaGluLeuAlaArgGlnGlnGluAlaGluArg 1328
3314 TTATGCGCGCGAGGAAGAGAAAAACGGTGCAGCGGATAAA 3357
|||::: |||::: |||::: |||::: |||::: |||
1329GlnAlaAlaGluLeuSerAlaLysGlnLysAlaGluThrAspArg 1343
3358 GACAGCGCTTGGGAAACAGCGCGGAAACCGCGCGCGGNTACCAC 3407
|||::: |||::: |||::: |||::: |||::: |||
1344 GluAlaAlaGluSerAlaLysArgLysAlaGluGluGluHisArgG 1360
3408 CGCCTTC 3414
|||
1360 nAlaAlaGlnSerGlnProGlnArgArgLysArgAlaAlaProGlnA 1377
3415CCCCGCGCGCGCGCGCGCGCGCGCGG 3438
|||::: |||
1377 spTyrMetAlaAlaSerGlnAsnArgProLysArgGlyArgArgSer 1393
3439 GATTTCGCGCAACCG 3453
|||::: |||
1394 ThrLeuProAlaProSerProSerPheAspSerSerAlaTyrAlaAl 1410
3453 3453
1410 aProArgAlaLeuHisAsnProAspTrpTyrGluAsnAspTyrGluGlu 1427
3453 3453
1427 leProLeuAspAlaLeuGluAspGluAsnValSerGluSerValAspThr 1443
3454CAGCCCCAA 3462
|||::: |||
1444 SerAspLysGlnProGlnAspAsnThrGluLeuHisGluLysTyrGluAs 1460
3462 3462
1460 nAspTyrGluGluIleProLeuAspAlaLeuGluAspGluAspValSerG 1477
3463CCGCAACCTCAA 3474
|||::: |||
1477 luserValAspThrSerAspLysGlnProGlnAspAsnThrGluLeuHis 1493
3475CCCCAACCG 3483
|||
1494 GluLysValGluThrValSerLeuGlnProArgAlaAlaGlnProArgAl 1510
3484CAGCGCGACCTGA 3496

101 LeuaspValSerAsnAspGluAsnGluTyrArgSerValAlaGlnAsnAs 117
435 TTATAAGCCTGACAAATCA.....CACCCCTTACAACGGC.....G 469
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 117
pTyrGluProAsnLysAsnTrpHisHisGlyAsnGlnGlyArgLeuGluA 134
470 ATTANCATATCCCGGTTTGGTATTAATTTGTCACAGATGACAGACCTGTC 519
|| :|||:||||:||||:||||:||||:||||:||||:||||:||||: 117
sPtyrAsnMetAlaArgLeuAsnLysPheValThrGluValAlaProile 150
520 GAAATGACGAGTACATGAGGGG...AATACCTATTCCGATAAAGAAA 566
|||||:||||:||||:||||:||||:||||:||||:||||:||||: 117
151 AlaProThrSerAlaGlyGlyValGluThrTyrLysAspLysAsnAr 167
567 ATATCCCGAGGTGTCCGATCGCTCAGGACACCACTATTGGGTATG 616
:|||:||||:||||:||||:||||:||||:||||:||||:||||: 117
167 gPheSerGluPheValArgValGlyAlaGlyThrGlnPheGluTyrAsnS 184
617 ATGATGACAAACAGCGGATTTATCTTACCTACCTCCGCGCATGTTAATGGC 666
|||: :|||:||||:||||:||||:||||:||||:||||:||||: 117
184 erArgTyrAsnMetThrGluLeuSerArgAlaTyrArgTyrAlaAla 200
667 GGCAATACACATATGCAG..... 684
201 GlyThrProTyrGlnAspValAsnValThrSerAsnLeuAsnGlnGlu 217
685GGTTGGGGAATAATAGCGGTANTTAGTTTGGCGCGCATGTC 727
:||||:||||:||||:||||:||||:||||:||||:||||:||||: 117
217 yLeuileGlyPheGlyAspAsnSerLysHisHisSerProGluLysLeu 234
728 GCCATGCCAAGCATATGGCCCTATGGCGATGGAGTTGAGGTGGCGGAC 777
:||||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: 117
234 ysGluValLeuSerGlnAsnAlaLeuThrAsnTyrAlaValLeuGlyAsp 250
778 AGCGTTCGCCAATGTTTATTATGACAAACAAACAAATTAATGGCT 827
|||:||||:||||:||||:||||:||||:||||:||||:||||: 117
251 SerGlySerProLeuPheAlaTyrAspLysGlnGluLysArgTrpValPh 267
828 CAACGGAGTTTACAA.....ACCGGTACCCCTTATTCGGCGAGGAAA 871
: |||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: 117
267 eLeuGlyAlaTyrAspTyrTrpAlaGlyTyr.....GlnLysA 280
872 ACGGTTTCCAG.....CTGATAGCGAAAGATTTGGTCTACGATGACATT 915
||||:||||: ||| ||| ||| ||| ||| ||| ||| ||| 117
280 snSerTrpGlnTrpAsnIleTyrLysLysGluPheAlaAspLysile 296
916 ...TACAGAGCGGATACACATACCGTCTNTTTTGAACCGCGCAGTACCG 962
|||:||||:||||:||||:||||:||||:||||:||||:||||: 117
297 LysGlnArgAspAsnAlaGlyThrile.....LysGlyAsnG 309
963 A.....CATTTTCTTTTACATCCAAACAAACGCGTACGGTACGGTAA 1006
| |||:||||:||||:||||:||||:||||:||||:||||:||||: 117
309 yGluHisHisTrpAsnIleThrPheGly..... 318
1007 CAGAACCAACGAAAGTNTCCATCCAAAGCTTAAGTACAGACATC 1056
||||:||||:||||:||||:||||:||||:||||:||||:||||: 117
319ThrAsnSerLysIleGlySer.....ThrAlaVal 328
1057 CGACTGTTTGAAGTCTTTGAATGAAGTAAAGTAAAGAACACCGTTTACG 1106
|||||:||||:||||:||||:||||:||||:||||:||||:||||: 117
329 ArgLeu.....AlaGlyAsnGluLysAsp..... 336
1107 GGCAGGGGGTGTATATCAGTACCGTCCCAAGGTTTAAACACGGTAAAC 1156
|||||:||||:||||:||||:||||:||||:||||:||||:||||: 117
337AlaAsnAsnGlyGlnAsnV 343
1157 TTTCTTTTATCGATTAGCGCAACGGCAACTCTTATCAACACATC 1206
:||||: ||| ||| ||| ||| ||| ||| ||| ||| ||| 117
343 alThrPheGluAsp.....AsnGlyThrLeuValLeuAspGlnAsnile 357
1207 AACCAAGCGCGGGGTTTGTATTTGAAGTGTATTTACGGTCTCGCC 1256
|||||:||||:||||:||||:||||:||||:||||:||||:||||: 117
358 AsnGlnGlyAlaGlyLeuPhePheLysGlyAspTyrThrValLysG 374

1257 TGAACAACAACGAA...ACGTGCAAGGCGCGGCGGCTTCATATCAGTGAAG 1303
|||||:||||:||||:||||:||||:||||:||||:||||:||||: 117
374 ylleAsnAsnAspIleThrTrpLeuGlyAlaGlylleAspValThrAspG 391
1304 ACAGTACCCTTACTTGGAAAGTAAACGGCGTGGCAACGACCGCTGTCC 1353
:|: |||:||||:||||:||||:||||:||||:||||:||||:||||: 117
391 lyLysLysValValTrpGlnValLysAsnProAsnGlyAspArgLeuAla 407
1354 AAAATCGCAAGGACGCTGACGCTTCAAGCAAGGCAAGGCAAGCAAG 1403
|||||:||||:||||:||||:||||:||||:||||:||||:||||: 117
408 LysIleGlyLysGlyThrLeuGluIleAsnGlyThrGlyValAsnGlnG 424
1404 CTCGATCAGCGTGGCGACGGTACAGTCAATTTTGGATCAGCAGGACAG 1453
:||||:||||:||||:||||:||||:||||:||||:||||:||||: 117
424 yGlnLeuLysValGlyAspGlyThrValIleLeuAsnGlnGlnAlaAsp 441
1454 ATAAAGCGCAAAACAAAGCCTTCTAGTGAATCGGCTTGTTCAGCGGAC 1503
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| 117
441 laAspLysLysValGlnAlaPheSerGlnValGlylleValSerGlyArg 457
1504 GGTACGGTGCACATGAATCCGATATCAAGTTCACCCGCAACACTCTA 1553
|||||:||||:||||:||||:||||:||||:||||:||||:||||: 117
458 GlyThrLeuValLeuAsnSerSerAsnGlnIleAsnProAspAsnLeuTy 474
1554 TTTGGCTTTTCGGCGGACGTTTGGATTAAACGGGATTCGCTTTCGT 1603
|||||:||||:||||:||||:||||:||||:||||:||||:||||: 117
474 rPheGlyPheArgGlyGlyArgLeuAspAlaAsnGlyAsnAspLeuThrP 491
1604 TCCACCGTATTCAAAATACCGATGAAGGGCGATGATTCGNCATCATAT 1653
:||||:||||:||||:||||:||||:||||:||||:||||:||||: 117
491 heGluHisIleArgAsnValAspGluGlyAlaArgIleValAsnHisAsn 507
1654 GCACAAACACATCCACCGTTACCATTCAGGGAATGAAGTATTACACA 1703
:||||: :|||:||||:||||:||||:||||:||||:||||:||||: 117
508 ThrSerHisAlaSerThrIleThrLeuThrGlyLysSerLeuIleThrAs 524
1704 ACCG..... 1707
: ||| 117
524 nProAsnSerLeuSerValHisSerIleGlnAsnAspTyrAspGluAspA 541
1707 1707
541 sPtyrSerTyrTyrTrpArgProArgProIleProGlnGlyLysAsp 557
1708AGTGGTAAAGAAATAT 1721
: |||: |||: |||: |||: |||: |||: |||: |||: |||: 117
558 LeuTyrTyrLysAsnTyrArgTyrTyrAlaLeuLysSerGlySerVa 574
1722 CAATAGA..... 1728
: ||| 117
574 laAsnAlaProMetProGluAsnGlyValThrGluAsnAsnAspTrpValP 591
1729 ..CTTAATTACAGCAAGAAATTCG..... 1752
: |||: |||: |||: |||: |||: |||: |||: |||: |||: 117
591 heMetGlyTyrThrGlnGluGluAlaLysLysAsnAlaMetAsnHisLys 607
1753TACAACGGTTCGTTTGGCGAGAAAGATAC 1781
: |||: |||: |||: |||: |||: |||: |||: |||: |||: 117
608 AsnAsnGlnArgIleSerGlyPheSerGlyPheGlyGluGluAsnG 624
1782 GACCAAAACGACGGCGGCTCAACCTTGTTCACGCCCGCCGAGAG 1831
: |||: |||: |||: |||: |||: |||: |||: |||: |||: 117
624 yLysGlyHisAsnGlyAlaLeuAsnLeuAsnPheAsnGlyLysSerAlaG 641
1832 ACCGACCCCGCTGCTTCCCGCGGCAACAAATTTAAACGGCAACATCAG 1881
: |||: |||: |||: |||: |||: |||: |||: |||: |||: 117
641 laAsnArgPheLeuLeuThrGlyThrAsnLeuAsnGlyLysIleSer 657
1882 CAAACAAACGCAACTGTTTTCAGCGCAGACCGACACCGACCGCCTA 1931
: |||: |||: |||: |||: |||: |||: |||: |||: |||: 117
658 ValThrGlnGlyAsnValLeuLeuSerGlyArgProThrProHisAlaAr 674

```

1932 CAATCATTTAGGAGCGGGTGTCAAAAATGGAAGGT.....ATCCAC 1975
...
674 gAspPheValAsnLysSerSerAlaArgLysAspAlaHisPheSerLysA 691
1976 AAGGAGAAATCGTGTGGGCAACAGCTGGATGCNACCGCAGCTTTAAAGCG 2025
...
691 snAsnGluValPheGluAspTrpIleAsnArgThrPheLysAla 707
2026 GAAATTTCCATATT...CAGGCGGCGCAGCGGTGTATTCCCGCAATGT 2072
...
708 AlaGluIleAlaValAsnGlnSerAlaSerPheSerSerGlyArgAsnVa 724
2073 TGCCAAAGTGAAGCGGATTTGACCAATTCACGCCCAAGCAGCTTT 2122
...
724 lSerAsnIleThrAlaAsnIleThrAlaThrAspAsnAlaLysValAsnL 741
2123 TTGGTGTCCGACCGCATCAACAGCATACAACTGTGTACAGTTCGGACTGG 2172
...
741 euGlyTyr.....LysAsnGlyAspGluValCysValArgSerAspTyr 755
2173 ACNGGTCTGACAAATTTGTGCGAANAANCATTACCGGACGATAAAGTGAT 2222
...
756 ThrGlyTyrValThrCys..... 761
2223 TGCTTCATTGACTAGACGACNTNAGCGGCANTGTNAGNCTNNCCNATN 2272
...
761 ..... 761
2273 ACGTNTNTNAANCNTCNGGCGNTGCNNCACTNAAGCAATCTTAGT 2322
...
762 .....AsnThrGlyAsnLeuSer 767
2323 GCAATGGCGGATACAGCTTATACAGTCAGCCACAGCCACCAA...AA 2369
...
768 .....AspLysAlaLeuAsnSerPheGlyAlaThrGlnIleAs 780
2370 CGGCACTTAGCTCGTGGGCAATGCCAAGCAACATTAATCAAGCCA 2419
...
780 nGlyAsnValAsnLeuAsnGlnAsnAlaLeuValLeuGlyLysAlaA 797
2420 CATTAAACGGCAACNATCGNTTCGGCAATCTTCATTAATCAAGC 2469
...
797 lAlaLeuTrpGlyGlnIleGlnGlyAsnSerArgValSerLeuAsn 813
2470 AACACGCGCCACAAAACGCGAGCTGTGAGCGTTTCGACAAAGCTAAGGC 2519
...
814 GlnHis..... 815
2520 AACGTAAGCCATTCCGACTCAACGGCAATGTCTCCCTAGCGGATAAGG 2569
...
815 ..... 815
2570 CAGTATTTCCATTTTGAACACAGCGCTTTACCGGACAACTACGCGCAGC 2619
...
815 ..... 815
2620 AAGGANACAGCATTTACACTTAAAGACAGCAATGAGCGCTGCGCTCAGC 2669
...
816 .....SerLysTrpHisLeuThrGlyAs 823
2670 CAGGGAATAGGCAATTTAACTTGCACAAAGCCACCATTAACACTCAAT 2719
...
823 pSerGlnValHisAsnLeuSerLeuAlaAspSerHisIleHisLeuAsnA 840
2720 CCGCTTATCGCCACGATGTGCGAGCGCGCAACCGGAGNGTGTGAGAC 2769
...
840 snAla.....SerAspAlaGlnSerAla..... 847
2770 ACGCCGCGCGCGTTCGCGCGCTTCCCTTATTATTCGTTACACGGCAAC 2819
...
847 ..... 847
2820 TTCGGTAGAATCCCGTTTCAACACGCTGACGGTAAACGGCAAAATTGAACN 2869

```

```

848 .....AsnGlnTyrHisThrLeuLysIleAsn...HisLeuSerG 860
2870 GTCAAGGAACATTCCTGCTTTATGTGGAACACTCTTCGGCTACCGAAGCAG 2919
...
860 lYAsnGlyHisPheHisTyrLeuThrHisLeuAlaGluAsnLeuGlyAsp 876
2920 AAATTGAAGCTGGCGGAAAGTTCCGAAGNACTTACACCTTGGCGGTCAA 2969
...
877 LysValLeuValLysGluSerAlaSerGlyHisTyrGlnLeuHisValGI 893
2970 CAATACCGCGCAACGACCC.....GTAAGCTCGATCAATGA 3007
...
893 nAspLysThrGlyGluProAsnGlnGluGlyLeuAsnLeuPheAspAlaAs 910
3008 CGGTAGTGAAGGAAAGAACACAAACCGCTGTCGAAACCTTAATTC 3057
...
910 erSerValGlnAspArg.....SerArgLeuSerVal 920
3058 ACCCTGCAAAAGCAACGTCGATCGCGCGCGCTTACCACTCAT 3107
...
921 SerLeuAlaAsnAsnHisValAspLeuGlyAlaLeuArgTyrThrIleLy 937
3108 CGCGAAGACGCGAGTTCCGCTCATATCCGCTCAAGCAAGCAAGAGC 3157
...
937 sThrGluAsnGlyIleThrArgLeuTyrAsnProTyrAlaGluAsnArgA 954
3158 TTTCGCAAACTCGGCAAGCAGAGCC.....AAAAACAGCGC 3198
...
954 rgArgValLysProValProSerProAlaThrAsnThrAlaSerGlnAla 970
3199 GAAAAAGACAAACGCGCAA 3216
...
971 GlnThrAspSerAlaGln 976
seq_name: sp_bacteria:09S6X2
seq_documentation_block:
ID Q9S6X2 PRELIMINARY; PRT; 997 AA.
AC Q9S6X2;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE IGAL1 PROTEASE (FRAGMENT).
GN IGA.
OS Neisseria meningitidis
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=487;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=24081;
RX MEDLINE=98010345; PubMed=9350862;
RA Morelli G., Malorny B., Muller K., Sellaer A., Wang J.F., del Valle J.,
RA Achtman M.;
RT "Clonal descent and microevolution of Neisseria meningitidis during 30
RT years of epidemic spread.";
RL Mol. Microbiol. 25:1047-1064(1997).
DR EMBL: AF012209; AAC45792.2;
DR MEROPS: S06.001;
DR InterPro: IPR002195; Dihydroorotase.
DR InterPro: IPR000710; Iga_S6.
DR Pfam: PF02395; IGAL1.
DR PRINTS: PR00921; IGASERPTASE.
DR PROSITE: PS00482; DIHYDROOROTASE_1; UNKNOWN_1.
KW Protease.
FT NON_TER 1
FT NON_TER 997
SQ SEQUENCE 997 AA; 109812 MW; 06F2E361E7E202E0 CRC64;

```

alignment_scores:

Quality: 1313.50

Length: 1160

Ratio: 2.021

Gaps: 34

Percent Similarity: 56.034 Percent Identity: 30.517

alignment_block:

US-09-303-518D-651 x Q9S6X2

Align seg 1/1 to: Q9S6X2 from: 1 to: 997

```
94 CTGTCGTCGCAATCTCCCAAGCTTGGCGGACACACATTTATTCGG 143
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1 LeuAlaTyrAlaLeuThrProTyrSerGluAlaAlaLeuValArgAspAs 17
144 CATCACTACCAATACATATCGGACATTTGGCGAAATAAAGGCAAGTTG 193
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
17 pValAspTyrGlnIlePheArgAspPheAlaGluAsnLysGlyLysPheP 34
194 CAGTCGGCGGCAAGATATTGAGGTNTACACAAACAAAGGGAGTTGGTC 243
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
34 heValGlyAlaThrAspLeuSerValLysAsnLysGlnGlyGlnAsnIle 50
244 GGCAATCAATGACAAAGCCCGATGATTGATTTCTGTGTCGCGCG 293
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
51 GlyAsnAlaLeuSerAsnValPrometIleAspPheSerValAlaAspVa 67
294 TAAC...GGCGTCGCGCATTTGGTGGCGATCAATATTTGTCAGCGTGG 340
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
67 lAsnArgArgThrLeuThrValIleAspProGlnTyrAlaValSerVal 84
341 CACATAACGGCGGTATTAACACGTTGATTTGGTCGGAAGGAAGNAAT 390
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
84 yShIsValLysGlyAspGluIleSerTyrTyrGlyHisHisAsnGlyHis 100
391 CCCGAT.....CAGCACCGCTTTTCTTACCAATTTGTGAAAGAAATAA 434
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
101 LeuAspValSerAsnAspGluAsnGluTyrArgSerValAlaGlnAsnAs 117
435 TTATAACCTGCACAAATCA.....CACCTTACACCGC.....G 469
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
117 pTyrGluProAsnLysAsnTrpHisHisGlyAsnGlnGlyArgLeuGlu 134
470 ATTANCATATGCCCGTTTTCGATAAATTTGTACAGATGCAGAACCTGTC 519
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
134 spTyrAsnMetAlaArgLeuAsnLysPheValThrGluValAlaProIle 150
520 GAATGACGAGTACATGAGGGGG...AATACCTATTCCGATAAAGAAA 566
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
151 AlaProThrSerAlaGlyGlyValGluThrTyrLysAspLysAsnAr 167
567 ATATCCCGCAGCGTCTCCGATCGCTCAGCACACCACTATTGCGGTATG 616
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
167 gPheSerGluPheValArgValGlyAlaGlyThrGlnPheGluTyrAsnS 184
617 ATGATGACAAACACGCGGATTTATCTTACCTCCGCGCATGTTAAATGGC 666
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
184 erArgTyrAsnMetThrGluLeuSerArgAlaTyrArgTyrAlaIleAla 200
667 GGCAATACATATGACAG.....ACCGCTACCTTATTCCGCGAGGAAA 871
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
201 GlyThrProTyrGlnAspValAsnValThrSerAsnLeuAsnGlnGlu 217
685 .....GGTTGGGAATAATAGCGGTANTTAGTTTCAGCGCGCATGTC 727
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
217 yLeuIleGlyPheGlyAspAsnSerLysHisHisSerProGluLysLeu 234
728 GCCATGCCACGAGTATGCGGCTATCCGATGTCAGGTGCGCGAGCGCAC 777
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
234 ysGluValLeuSerGlnAsnAlaLeuThrAsnTyrAlaValLeuGlyAsp 250
778 AGCGGTCGCAATGTTTATTTATGACAAACAAACAAATAAATGCGTGTCT 827
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
251 SerGlySerProLeuPheAlaTyrAspLysGlnGluLysArgTrpValPh 267
828 CAACGAGGTTTACAA.....ACCGCTACCTTATTCCGCGAGGAAA 871
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

```
267 eLeuGlyAlaTyrAspTyrTrpAlaGlyTyr.....GlnLysA 280
872 ACGGTTTTCAG.....CTGATACGCAAAAGATTGGTTCTTACGATGACATT 915
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
280 snSerTrpGlnGluTrpAsnIleTyrLysGluPheAlaAspLysIle 296
916 ...TACAGAGGCGATACATACCGTCTNTTTTGAACCGCGCATACGG 962
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
297 LysGlnArgAspAsnAlaGlyThrIle.....LysGlyAsnG1 309
963 A.....CATTTTCTTACATCCACACACACACGATACGGGTACGGTAA 1006
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
309 yGluHisHisTrpAsnIleThrPheGly..... 318
1007 CAGAAACCAACGAAAGGTTNTCCAATCCAAAGCTTAAAGTACAGACAGTC 1056
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
319 .....ThrAsnSerLysIleGlySer.....ThrAlaVal 328
1057 CGACGTGTTTGAGCAATCTTTGAATGAAACTGATAAAGAACAGTTTACGC 1106
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
329 ArgLeu.....AlaGlyAsnGluLysAsp..... 336
1107 GCGAGGGGTGTTTAACTCAGTACCGTCCAAAGTTTAAACACGTTGAAAC 1156
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
337 .....AlaAsnAsnGlyGlnAsnV 343
1157 TTTCTTTTATCGATTACGCAACGCAAACTCATCTTATCAAAACAAATC 1206
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
343 alThrPheGluAsp.....AsnGlyThrLeuValLeuAspGlnAsnIle 357
1207 AACCAAGCGCGGCGGTTTGTATTTTGAAGTGATTTTACGGTCTCGCC 1256
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
358 AsnGlnGlyAlaGlyGlyLeuPhePheLysGlyAspTyrThrValLysG1 374
1257 TCAAAACACACGAA...ACGTGCAAGCGCGCGCTTCATCATCAGTGAAG 1303
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
374 yIleAsnAsnAspIleThrTrpLeuGlyAlaGlyIleAspValThrAspG 391
1304 ACAGTACCGTTACTTGGAAAGTAAACGCGTGGCAACGACCGCTGTCC 1353
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
391 yLysLysValValTrpGlnValLysAsnProAsnGlyAspArgLeuAla 407
1354 AAAATCGGCAAGGACCGTGCACGTTCAAGCCAAAGGGGAAACCAAGG 1403
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
408 LysIleGlyLysGlyThrLeuGluIleAsnGlyThrGlyValAsnGlnG1 424
1404 CTCGATCAGCGTGGCGGACGTTACAGTCACTTTTGGATCAGCAGGACAG 1453
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
424 yGlnLeuLysValGlyAspGlyThrValLeuAsnGlnGlnAlaAspA 441
1454 ATAAAGGCAAAACACGCTTTTACTGAAATCGGCTTGTTCAGCGGACAG 1503
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
441 laAspLysLysValGlnAlaPheSerGlnValGlyIleValSerGlyArg 457
1504 GGTACGGTGCACACTGAATCCCGATATCAGTTCAACCCCGACAACTCTA 1553
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
458 GlyThrLeuValLeuAsnSerSerAsnGlnIleAsnProAspAsnLeu 474
1554 TTTTCGGCTTTCGCGGCGGACGTTTGGATTAAACGGGCAATTCGCTTTCGT 1603
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
474 rPheGlyPheArgGlyGlyArgLeuAspAlaAsnGlyAsnAspLeuThrP 491
1604 TCCACCGCTATTCAAAATACCGATGAAGGGCGGATGATTGNCNATCAAT 1653
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
491 heGluHisIleArgAsnValAspGluGlyAlaArgIleValAsnHisAsn 507
1654 GCCAACACACATCCCGTTTACCATTCAGCGAATGAAGTATTATACACA 1703
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
508 ThrSerHisAlaSerThrIleThrLeuThrGlyLysSerLeuIleThrAs 524
1704 ACCG..... 1707
|||||:
524 nProAsnSerLeuSerValHisSerIleGlnAsnAspTyrAspGluAspA 541
```



```
1707 ..... 1707
541 spTyrSerTyrTyrArgProArgProIleProGlnGlyLysAsp 557
1708 .....AGTGGTAAGATAT 1721
558 LeuTyrTyrLysAsnTyrArgTyrTyrAlaLeuLysSerGlyGlySerVa 574
1722 CAATAGA..... 1728
574 IAsnAlaProMetProGluAsnGlyValThrGluAsnAsnTyrPValP 591
1729 ..CTTAATTACGACAAAGAAATTGCC..... 1752
591 heMetGlyTyrThrGlnGluAlaLysLysAsnAlaMetAsnHisLys 607
1753 .....TCAACGGTGGTTGGCGAGAAATAC 1781
608 AsnAsnGlnArgIleSerGlyPheSerGlyPheGlyGluGluAsnG1 624
1782 GACCAAAACGACGGCGCTCAACCTGTGTACCGACCGCCGCGCAGAAG 1831
624 YLysGlyHisAsnGlyAlaLeuAsnLeuAsnPheAsnGlyLysSerAlaG 641
1832 ACCGACCCNGCTGCTTTCGGCGGGAACAAATTTAAACGGCAACATCAGC 1881
641 InAsnArgPheLeuLeuThrGlyThrAsnLeuAsnGlyLysIleSer 657
1882 CAACAAACGGCAACTGTTTTTCAGCGGACGACCGACCGCAGCCCTA 1931
658 ValThrGlnGlyAsnValLeuLeuSerGlyArgProThrProHisAlaAr 674
1932 CAATCATTTAGGAAGCGGTGGTCAAAATGGAAGT.....ATCCAC 1975
674 gAspPheValAsnLysSerSerAlaArgLysAspAlaHisPheSerLysA 691
1976 AAGGAGAAATCGTGTGGCAACAGCTGAGTCNACCGCAGCTTTAAAGCG 2025
691 snAsnGluValValPheGluAspAspTrpIleAsnArgThrPheLysAla 707
2026 GAAATTTCCATATT...CAGGCGGCGCAGCGGTGATTTCGCGCAATGT 2072
708 AlaGluIleAlaValAsnGlnSerAlaSerPheSerSerGlyArgAsnVa 724
2073 TGCCAAAGTGAAGCGATTGNCATTTGACCAATCAGCCCAAGCAGTTT 2122
724 lSerAsnIleThrAlaAsnIleThrAlaThrAspAsnAlaLysValAsnL 741
2123 TTGGTGTGCGCACCGCATCAAGCATACATACTGTACAGCTTCGGACTGG 2172
741 euGlyTyr.....LysAsnGlyAspGluValCysValArgSerAspTyr 755
2173 ACNGGTCGCAAAATTGTGTGCAANAANAANCATTACCGACGATAAAGTAT 2222
756 ThrGlyTyrValThrCys..... 761
2223 TGCTTCATTGACTAAGACACACNTNACCGGCANTGTNAGCTNNCCNATN 2272
761 ..... 761
2273 ACGTNTNTTNAANCTCNCGGCGTGCNNCACTNAAGGCAATCTTAGT 2322
762 .....AsnThrGlyAsnLeuSer 767
2323 GCAATGGCGATACAGTTATACAGTCAGCCACACGCCACCCAA...AA 2369
768 .....AspLysAlaLeuAsnSerPheGlyAlaThrGlnIleAs 780
2370 CGGCAACTTAGCTCTGCGGCAATGCCCAAGCAACATTTAATCAAGCCA 2419
780 nGlyAsnValAsnLeuAsnGlnAsnAlaAlaLeuValLeuGlyLysAlaA 797
```

seq_name: sp_bacteria:Q57309

seq_documentation_block:

ID Q57309 PRELIMINARY;

AC Q57309;

PRT;

996 AA.

[illegible]

```

DR MEROPS; S06.001; -.
DR InterPro; IPR002195; Dihydroorotase.
DR InterPro; IPR000710; Iga_S6.
DR Pfam; PF02395; IGAL; 1.
DR PRINTS; PR00921; IGASERPTASE.
DR PROSITE; PS00482; DIHYDROOROTASE_1; UNKNOWN_1.
KW Protease..
FT NON_TER 1
FT NON_TER 992
FT SEQUENCE 992 AA; 109228 MW; 3677DDE4CE6D9F69 CRC64;

alignment_scores:
  Quality: 1309.00      Length: 1149
  Ratio: 2.020          Gaps: 31
Percent Similarity: 56.397      Percent Identity: 30.548

alignment_block:
US-09-303-S18D-651 x Q57035  ..

Align seg 1/1 to: Q57035 from: 1 to: 992

94 CTGTCGTTCGGCATCTCTCCCAAGCTTGGCGGGACACACACTTATTTTCGG 143
1 LeuAlaTyrAlaLeuThrProTyrSerGluAlaLeuValArgAspAs 17

144 CATCAACTACCAATACTATCGCGACTTCCGCAAAATAAGCAAGTTTG 193
pValAspTyrGlnIlePheArgAspPheAlaGluAsnLysGlyLysPheP 34

194 CAGTCGGGGGAAGATATTGAGTGTWACACAAAAGGGAGGTGGTC 243
heValGlyAlaThrAspLeuSerValLysAsnLysGlnGlyGlnAsnIle 50

244 GGCAAAATCAATGACAAAACCCCGATGATTGTTCTGTCGTGTCGCG 293
GlyAsnAlaLeuSerAsnValProMetIleAspPheSerValAlaAspVa 67

294 TAAC...GGCGTGGCGGCATGTGTGGCGCATCAATATATTGTGACGCTGG 340
lAsnArgArgThrLeuThrValIleAspProGlnTyrAlaValSerValL 84

341 CACATAACGGCGCTATACACAGCTGTTGGTGGCGGAAGAAGNAAT 390
ysHisValLysGlyAspGluIleSerTyrTyrGlyHisAsnGlyHis 100

391 CCCGAT.....CAGCACCGCTTTCTTACCAGAAATGTGCAAGAAATAA 434
LeuAspValSerAsnAspGluAsnGluTyrArgSerValAlaGlnAsnAs 117

435 TTATAAGCTTGACAAATCA.....CACCTTACACAGGC.....G 469
pTyrGluProAsnLysAsnTrpHisHisGlyAsnGlnGlyArgLeuGluA 134

470 ATTANCAATTCGCCGCTTTCGCATAAAATTTGTCACAGATGCAGAACCTGTC 519
sPTyrAsnMetAlaArgLeuAsnLysPheValThrGluValAlaProIle 150

520 GAAATGACGAGTGCATGAGGGGG...AATACCTATTCCGATAAAGAAA 566
AlaProThrSerAlaGlyGlyValGluThrTyrLysAspLysAsnArg 167

567 ATATCCGCGCGGTCCGATCGGCTCAGACACCACTATTGGCGTTATG 616
gPheSerGluPheValArgValGlyAlaGlyThrGlnPheGluTyrAsnS 184

617 ATGATGACAAACACGCGCATTTTCTACTCCGGCGCATGTGTAATTGGC 666
erArgTyrAsnMetThrGluLeuSerArgAlaTyrArgTyrAlaIleAla 200

667 GGCAATACACATGTCAG..... 684

201 GlyThrProTyrGlnAspValAsnValThrSerAsnLeuAsnGlnGluGl 217

```

1563 TCAGCGCGGCGCTTTGGATTTAAACGGGCATTCGCTTTCGTTCACCGTA 161
161 eArgGlyGlyPheGlyAspAsnSerLysHisSerProGluLysLeuL 234
162 yLeuIleGlyPheGlyAspAsnSerLysHisSerProGluLysLeuL 234
163 GCATGCCAAGCACTATGCGCCCTATCCGATTCGAGTTCGCGCAGCGAC 777
728 yGluValLeuSerGlnAsnAlaLeuThrAsnTyrAlaValLeuGlyAsp 250
234 yGluValLeuSerGlnAsnAlaLeuThrAsnTyrAlaValLeuGlyAsp 250
778 AGCGGTTCGCAATGTTTATTTATGACAAACAAACAAATAAATGCTGCT 827
779 yGluValLeuSerGlnAsnAlaLeuThrAsnTyrAlaValLeuGlyAsp 250
251 SerGlySerProLeuPheAlaTyrAspLysGlnLysGlyArgTyrValPh 267
252 yGluValLeuSerGlnAsnAlaLeuThrAsnTyrAlaValLeuGlyAsp 250
828 CAACGGAGTTTACAA.....ACCGCTACCCCTTATTCGCGCAGGAAA 871
267 eLeuGlyAlaTyrAspTyrTrpAlaGlyTyr.....GlnLysA 280
872 ACGGTTTCCAG.....CTGATACGCAAAAGATTGGTTTACGATGACATT 915
873 yGluValLeuSerGlnAsnAlaLeuThrAsnTyrAlaValLeuGlyAsp 250
280 snSerTrpGlnGluTrpAsnIleTyrLysLysGluPheAlaAspGluIle 296
297 yGluValLeuSerGlnAsnAlaLeuThrAsnTyrAlaValLeuGlyAsp 250
916 TACAGAGCGGATACATACCCGCTCTNTTTTGAACCGCGCAGTAAACGGACA 965
297 LysGlnArgAsp...AsnAlaGlyThrIleLysGlyTyrGlyGluHisH 312
966 TTTTTCCTTTACATCAACAAACAGCGTACGGGTACGGTAAACAGAAACCA 1015
312 strLysThrThrGlyThrAsnSerHisIleGlySerThr..... 325
1016 ACGAAAAGGNTCCAAATCCAAAGCTTAAAGTACAGACAGTCCGACTGTT 1065
326AlaValArgLeu... 329
1066 GAGCAATCTTTGAATGAACTGATTAAGAACCAAGTTTACGGCGAGGGG 1115
330AlaGlyAs 332
1116 TGTTAATCAGTACCGTCAAGGTTAAACAAACGGTGAACACCTTCTTTTA 1165
332 n.....GluArgGlyAlaAsnAsnGlnAsnValThrPhe. 344
1166 TCGATTACGCAACGGCAACATCATCTTATCAACAAACATCAACCAAGGC 1215
345GluAsnAsnGlyThrLeuValLeuAspGlnAsnIleAsnGlnGly 359
1216 GCGGCGGTTTCTATTTTCAAGGTGATTTACGGTCTCGCTGAAACAA 1265
360 AlaGlyGlyLeuPhePheLysGlyAspTyrThrValLysGlyIleAsnAs 376
1266 CGAA...ACGTGGCAAGCGCGGGCGTTCATATCAGTGAAGACAGTACCG 1312
376 nAspIleThrTrpLeuGlyAlaGlyIleAspValAlaAspGlyLysLysV 393
1313 TTTATGGAAGTAACGCGGTGGCAACAGCGCTCTCGCTGAAATCGGC 1362
393 alValTrpGlnValLysAsnProAsnGlnValAspArgLeuAlaLysIleGly 409
1363 AAAGCAGCGCTCAGCTTCAAGCCAAAGGGGAAACCAACAGCTCGATCAG 1412
410 LysGlyThrLeuGluIleAsnGlyThrGlyValAsnGlnGlyGlnLeuLy 426
1413 CTGGGCGCAGGTACAGTCATTTGGATTCAGCAGCGCAGCAGATAAAGCA 1462
426 sValGlyAspGlyThrValIleLeuAsnGlnLysAlaAspSerAsnGlnL 443
1463 AAAAACAAGCCCTTATGTAATTCGGCTTGNCTCAGCGCAGGAGTACGGTG 1512
443 yValGlnAlaPheSerGlnValGlyIleValSerGlyArgGlyThrLeu 459
1513 CAACGTAAATGCGGATAATACGTTCAACCCCGCAGCAAACTCTATTTTCGGCT 1562
460 ValLeuAsnSerSerAsnGlnIleAsnProAspAsnLeuTyrPheGlyPh 476

760 760
2282 NAAANCTCNMGCGTGCNNACCTNAANGCAATCTTAGTGCAATGGC 2331
761 766
2332 GATACACGTTATACAGTCAGCCACCAAGCCACCCAA...AACGGCAACCT 2378
767 ...AspLysAlaLeuAsnSerPheGlyAlaThrGlnIleAsnGlyAsnVa 782
2379 TAGCCTCGTGGCAATGCCCAAGCAATTAATCAAGCCACATTAACG 2428
782 lasnLeuSerGlnAsnAlaLeuThrLeuGlyLysAlaLeuLeuTrpG 799
2429 GCAACNCACTCGNTTCGGCAATGCTTCATTTAATCAAGCAACACGCC 2478
799 lyGlnIleGlnGlyGlnGlyAsnSerArgValSerLeuAsnGlnHis... 814
2479 GCACAAACGGCAGTCTGACGCTTTCGACAAACGCTAAGGCAACGTAAG 2528
814 814
2529 CCAATTCGCACTCAACGGCAATGTCTCCCTAGCCGATAAGCCAGTATTC 2578
814 814
2579 ATTTGAAACAGCGCTTTACCGGCAACCTCAGCGGCAGCAAGANACA 2628
814 814
2629 GCATTACACTTAAAGACAGCAATGAGCGCTGCGCTCAGCAGCGGAAT 2678
815 825
2679 AGGCAATTTAAACCTTGACAGCCACCACTTACACTCAATTCGCGCTATC 2728
825 lhisAsnLeuSerLeuAlaAspSerHisIleHisLeuAsnAla... 840
2729 GCCAGATCTCTCAGCGCGCAACCCGAGCTGCTCAGACACGCGCGC 2778
841 ...SerAspAlaGlnSerAla... 846
2779 CGCGCTTCGCGCGCTTCCCTATTATTCCTTACCGCAACTTCGGTAGA 2828
846 846
2829 ATCCCGTTTCAACAGCTGACGGTAACCGCAATTTGAACNGTCAAGAA 2878
847 .AsnLysTyrHisThrLeuLysIleAsn...HisLeuSerGlyAsnGlyH 862
2879 CATTCGCTTTATGTCGGAACCTTTCGGCTACCGAAGCGCAAAATTTGAAG 2928
862 isPheHisThrLeuThrHisLeuAlaLysAsnLeuGlyAspLysValLeu 878
2929 CTGGCGGAAAGTTCGGAAGNACTTTACACTTGGCGGTCAACAATPACGG 2978
879 ValLysGluSerAlaSerGlyHisTyrGlnLeuHisValGlnAspLysTh 895
2979 CAACGAAACCGTAAGCTCGATCAATTCAGCGTAGTGGAGGAAAGACA 3028
895 rGlyGluPro...AsnGlnGluGlyLeuAsnLeuPheAspAlaSerServ 911
3029 ACAACCGCTGTCCGAAACCTTTAATTTACCCCTGCAAAACGAAACACGTC 3078
911 alArgAspArgSer...HisLeuSerValSerLeuAlaAsnHisVal 926
3079 GATCGCGCGCGTGGGTGTACCACTCATTCGCAAGAGCGGCGGATTCGG 3128
927 AspLeuGlyAlaLeuArgTyrThrIleLysThrGluAsnGlyIleThrAr 943
3129 CTGGCATATTCGGTCAAGAAACAGAGCTTTCGACAAACCTCGCAAGG 3178
|||||

943 gLeuTyrAsnProTyrAlaGluAsnArgArgValLysProAlaProS 960
3179 CAGAGACC.....AAAAACAGCGCGGAAAAAGACAAACGCGCAA 3216
960 erProAlaThrAsnThrAlaSerGlnAlaGlnThrAspSerAlaGln 975
seq_name: sp_bacteria:O30575
seq_documentation_block:
ID O30575 PRELIMINARY; PRT; 997 AA.
AC O30575;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE IGAI PROTEASE (FRAGMENT).
OS Neisseria meningitidis.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=487;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=24099;
RX MEDLINE=98010345; PubMed=9350862;
RA Morelli G., Malorny B., Muller K., Seiler A., Wang J.F., del Valle J.,
RA Achman M.;
RT "Clonal descent and microevolution of Neisseria meningitidis during 30
RT years of epidemic spread";
RL Mol. Microbiol. 25:1047-1064(1997).
DR EMBL: AF012210; AAC45793.2; -;
DR MEROPS: S06.001; -;
DR InterPro: IPR002195; Dihydroorotase.
DR InterPro: IPR000710; IGA_S6.
DR Pfam: PF02395; IGAI; 1.
DR PRINTS: PR00921; IGASERPTASE.
DR PROSITE: PS00482; DIHYDROOROTASE_1; UNKNOWN_1.
KW Protease.
FT NON_TER 1 997
FT NON_TER 997 997
SQ SEQUENCE 997 AA; 110152 MW; 4DE40DE594E5E28E CRC64;
alignment_scores:
Quality: 1308.50 Length: 1153
Ratio: 2.007 Gaps: 32
Percent Similarity: 56.548 Percent Identity: 30.442
alignment_block:
US-09-303-518D-651 x O30575 ..
Align seg 1/1 to: O30575 from: 1 to: 997
94 CTGTGTCGCGCATTCCTCCCAAGCTTGGCGGCGACACACTATTATTCGG 143
|||||
1 LeuAlaTyrAlaLeuThrProTyrSerGluAlaLeuValArgAspAs 17
144 CATCACTACCACTATATCGCGACTTTCGCCGAAATATAGGCAAGTTG 193
|||||
17 pValAspTyrGlnIlePheArgAspPheAlaGluAsnLysGlyLysPhep 34
194 CAGTCGGCGGCGAAAGATATTGAGGTNTACACAAAGGAGGAGTTGGTC 243
|||||
34 heValGlyAlaThrAspLeuSerValLysAsnLysGlnGlyGlnAsnIle 50
244 GGCAAAATCAATGACAAACCCCGATGATTTCTGTGTGTGTCGGC 293
|||||
51 GlyAsnAlaLeuSerAsnValProMetIleAspPheSerValAlaaspVa 67
294 TAAC....GGCGTGGCGCATTTGGTGGCGATCAATATTATTGAGCGTGG 340
|||||
67 lAsnArgThrLeuThrValIleAspProGlnTyrAlaValSerVal 84
341 CACATAACGGCGGTATACACAGCTTGTGTCGGGAGGAAGNAAT 390
|||||

1707 1707
526 nLeuSerValTyrGluIleArgAsnAspTyrAspAspAspTyrTyrG 543
1707 1707
543 lYTyTyrSerTyrArgLysProIleProGlnGlyLysAspLeuTyrTyr 559
1708 AGTGGTAAGATATCATAGA... 1728
560 LysAsnTyrArgTyrTyrAlaLeuLysSerGlySerValAsnAlaPr 576
1729CTTAATT 1735
576 oMetProGluAsnGlyValThrGluAsnAsnAspTrpValPheMetGlyT 593
1736 ACAGCAAGAAATAGCC..... 1752
593 yrThrGlnGluGluAlaLysLysAsnAlaMetAsnHisLysAsnAsnGln 609
1753TACACGGTTGGTGGCGAGAAAGATAGCAGCAAAAC 1790
610 ArgIleSerGlyPheSerGlyPheGlyGluGluAsnGlyLysGlyHI 626
1791 GAACGGCGGCTCAACCTTGTGTACCGCCGCGCAGAGCAGCACCC 1840
626 sasnglyAlaLeuAsnLeuAsnPheAsnGlyLysSerAlaGlnAsnArgp 643
1841 NGCTGCTTCGGCGGAACAATTTAAACGGCAACATCAGCAGCAAAAC 1890
643 heLeuLeuThrGlyGlyAlaAsnLeuAsnGlyLysIleSerValThrGln 659
1891 GCGAACTGTTTTTCAGCGCGCAGACACCGCAGCGCTACATCATTT 1940
660 GlyAsnValLeuLeuSerGlyArgProThrProHisAlaArgAspPheVa 676
1941 AGGAAGCGGTGGTCAAAAATGGAAGT.....ATCCCAAGAGAGAAA 1984
676 lAsnLysSerSerAlaArgLysAspAlaHisPheSerLysAsnAsnGluV 693
1985 TCGTGTGGCAACGACGTGGATCNACCCGCGTAAAGCGGAAATTC 2034
693 alValPheGluAspSerTrpIleAsnArgThrPheLysAlaAlaGluile 709
2035 CATATT...CAGGGCGCGCAGCGGTGATTCCCGCAATGTGCAAGT 2081
710 ThrValAsnGlnSerAlaSerPheSerSerGlyArgAsnValSerAspII 726
2082 GGAAGCGATTGNCATTTAGCAATCAGCCCAAGCAGTGTGTGTGCG 2131
726 eThrAlaAsnIleThrAlaThrAspAsnAlaLysValAsnLeuGlyTyr. 742
2132 CACCCCATCAAGCATACATCTGTACACGTCGGACTGGACNGTCTG 2181
743LysAsnGlyAspGluValCysValArgSerAspTyrThrGlyTyr 757
2182 ACAAAATTGTGCGAANAANCATTACCGCAGTAAAGTGATTGCTTCATT 2231
758 ValThrCys..... 760
2232 GACTAAGACNACNTNACGGCGANTGTNAGNCTNNCCNATNACNTNNTT 2281
761AsnThrAsp..... 763
2282 NAAANCTCNCGGGGONTGCNCACTNNAAGCAATCTTAGTGCAATAGGC 2331
764AsnLeuSer..... 766
2332 GATACACGTTATACAGTCAGCCCAACGCCACCCAA...AACGGCAACCT 2378
767AspLysAlaLeuAsnSerPheGlyAlaThrGlnIleAsnGlyAsnVa 782

2379 TAGCCTCGTGGCAATGCCCAAGCAACATTTAATCAAGCCACATTAAACG 2428
782 lAsnLeuSerGlnAsnAlaAlaLeuValLeuGlyLysAlaAlaLeuTrpG 799
2429 GCAACNCATCGGNTTCGGCAATGCTTCAATTAATCTAAGCAACACGCC 2478
799 lylGlnIleGlnGlyAsnSerArgValSerLeuAsnGlnHis... 814
2479 GCACAAAACGGCAGTCTGACGCTTCCGACACACGCTAAGGCAACGTAAG 2528
814 814
2529 CCATTCCGCACCTCAAGGCAATGTCTCCCTAGCGATAGGCAAGTATCC 2578
814 814
2579 ATTTTGAACAGCCGCTTTACGGACACTCAGCGGCAGCAAGGANACA 2628
814 814
2629 GCATTACACTTAAAGACAGCAATGGACGCTCGCGTCAGGCACGGAATT 2678
815SerLysTrpHisLeuThrGlyAspSerGlnVa 825
2679 AGCAATTTAAACCTTGACACGCCACCATTTACATCAATTCGCCCTATC 2728
825 lHisAsnLeuSerLeuAlaAspSerHisIleHisLeuAsnAsnAla... 840
2729 GCCACATGCTCGAGCGCGCAACCGCAGNGTGTGACACACGCGCGC 2778
841 ..SerAspAlaGlnSerAla..... 846
2779 CGCGCTCGCGCGCTTATTCCTATTATTCCTGTACACCGCAACTTCGGTAGA 2828
846 846
2829 ATCCCTTTACACACGCTCAGGTAACGGCAATTTGACNCTCAAGNA 2878
847 .AsnLysTyrHisThrLeuLysIleAsn...HisLeuSerGlyAsnGlyH 862
2879 CATTCGCTTTATGCGGAACCTCTCGGCTACCGAAGCAGCAAAATTTGAAG 2928
862 isPheHisTyrLeuThrHisLeuAlaLysAsnLeuGlyAspLysValLeu 878
2929 CTGGCGGAAAGTTCGAAGNACTTACACCTTGGCGGTCAACAATACCGG 2978
879 ValLysGluSerAlaSerGlyHisTyrGlnLeuHisValGlnAspLysTh 895
2979 CAACGAAACC.....GTAGCCTCGATCAATTTGACGGTAGTGG 3016
895 rGlyGluProAsnGlnGluGlyLeuAsnLeuPheAspAlaSerSerValG 912
3017 AAGGAAAGACAAACACCGCTGTCGAAAAACCTTAATTTACCCCTGCAA 3066
912 lAspArg.....SerArgLeuSerValSerLeuAla 922
3067 AACGAACACGTCGATCCGCGCGCTTACCACTCATCCGCAAGA 3116
923 AsnAsnHisValAspLeuGlyAlaLeuArgTyrThrIleLysThrGluAs 939
3117 CGCGGAGTTCGCTCATTAATCCGTCAAAGCAAGAGCTTCCGACA 3166
939 nGlyIleThrArgLeuTyrAsnProTyrAlaGluAsnArgArgValL 956
3167 AACTCGGC.....AAGCAGAGCCAAAAAACAG 3195
956 ysProAlaProSerProAlaThrAsnThrAlaSerGlnAlaGlnLysAla 972
3196 GCGGAAAAAGACAAACCGCAA 3216
973 ThrGlnThrAspGlyAlaGln 979

seq_name: sp_bacteria.Q956X4

```

seq_documentation_block:
ID Q9S6X4 PRELIMINARY; PRT; 992 AA.
AC Q9S6X4;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE IGAL PROTEASE (FRAGMENT).
GN IGAL
OS Neisseria meningitidis.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=487;
RN [1]
RP
RC STRAIN=Z3906;
RX MEDLINE=98010345; PubMed=9350862;
RA Morelli G., Malorny B., Muller K., Seiler A., Wang J.F., del Valle J.,
RA Achman M.;
RT "Clonal descent and microevolution of Neisseria meningitidis during 30
RT years of epidemic spread.";
RL Mol. Microbiol. 25:1047-1064 (1997).
DR EMBL; AF012204; AAC45787.1;
DR MEROPS; S06.001;
DR InterPro; IPR002195; Dihydroorotase.
DR InterPro; IPR000710; IGA_S6.
DR Pfam; PF02395; IGAL1; 1.
DR PRINTS; PR00921; IGASERPTASE.
DR PROSITE; PS00482; DIHYDROOROTASE_1; UNKNOWN_1.
KW Protease.
FT NON_TER 1
FT NON_TER 992
SQ SEQUENCE 992 AA; 109239 MW; 9ED0A41019917CE3 CRC64;

alignment_scores:
Quality: 1285.00 Length: 1149
Ratio: 1.995 Gaps: 31
Percent Similarity: 56.049 Percent Identity: 30.287

alignment_block:
US-09-303-518D-651 x Q9S6X4
Align seg 1/1 to: Q9S6X4 from: 1 to: 992

94 CTGTCGTCGGCATCTTCCCAAGCTGGGGGACACACTTATTTCGG 143
|||||
1 LeuAlaTyAlaLeuThrProTyrSerGluAlaAlaLeuValArgAspAs 17
144 CATCACTACCAATCTACTCGCGACTTTGCCGAAATAAAGCAAGTTG 193
|||||
17 pValAspTyrGlnIlePheArgAspPheAlaGluAsnLysGlyLysPheP 34
194 CAGTCGGCGGCAAGATATTGAGTNTACACAAAAGGGAGGTGTC 243
|||||
34 heValGlyAlaThrLeuSerValLysAsnLysGlnGlyGlnAsnIle 50
244 GGCAATCAATGACAAAGCCCGCATGATGATTTTCTGTTGGTCCGG 293
|||||
51 GlyAsnAlaLeuSerAsnValProMetIleAspPheSerValAlaAspVa 67
294 TAAAC...GGCGTGGCGGATGGTGGCGGATCAATATATTGTGACGCTGG 340
|||||
67 IAsnArgArgThrLeuThrValIleAspProGlnTyrAlaValSerValL 84
341 CACATACGGCGCTATACACACTGTTGTTTGGTCGGGAAGAGNAAT 390
|||||
84 yHisValLysGlyAspGluLeuSerTyrTyrGlyHisAsnGlyHis 100
391 CCCGAT.....CAGCAGCGTTTCTTACCAAAATGTAAGAAAGAAATAA 434
|||||
101 LeuAspValSerAsnAspGluAsnGluTyrArgSerValAlaGlnAsnAs 117
435 TTATAAGCCTGACAAATTCACCCCTTACAACGGC.....G 469

```

```

|||||
117 pTyrGluProAsnLysAsnTrpHisHisGlyAsnGlnGlyArgLeuGluA 134
470 ATTANCATATCCGGTTCGATAAATTTGTACAGATCCAGAACCTGTC 519
|||
134 spTyrAsnMetAlaArgLeuAsnLysPheValThrGluValAlaProIle 150
520 GAAATGACAGTGACATGAGGGG...AATACCTATTCGATAAAGAAA 566
|||||
151 AlaProThrSerAlaGlyGlyValGluThrTyrLysAspLysAsnAr 167
567 ATATCCCGAGCTGTCGATCGCTCAGGACACCACTATTGGCCTTATG 616
|||||
167 gPheSerGluPheValArgValGlyAlaGlyThrGlnPheGluTyrAsnS 184
|||
184 erArgTyrAsnMetThrGluLeuSerArgAlaTyrArgTyrAlaIleAla 200
667 GGCAATACACATATGCAG..... 684
|||||
201 GlyThrProTyrGlnAspValAsnValThrSerAsnLeuAsnGlnGluL 217
685 .....GGTTGGGAAATATGCGGTANTTACGTTGAGCGCGCATGTGC 727
|||||
217 yLeuIleGlyPheGlyAspAsnSerLysHisHisSerProGluLysLeuL 234
728 GCCATGCCAAGCACTATGCGCCTATGCGGATTCGAGGTGCGCAGCGCAC 777
|||||
234 yGluValLeuSerGlnAsnAlaLeuThrAsnTyrAlaValLeuGlyAsp 250
778 AGCGGTTCGCAATGTTTATGACAAAACAACAATAATGCGTGCT 827
|||||
251 SerGlySerProLeuPheAlaTyrAspLysGlnGluLysArgTyrPValPh 267
828 CAACGGAGCTTTTACAA.....ACCGCTACCTTTATTCGGCAGGGA 871
|||||
267 eLeuGlyAlaTyrAspTyrTrpAlaGlyTyr.....GlnLysA 280
872 ACGTTTTCAG.....CTGATACGCAAGATTGGTTCTACGATGACATT 915
|||||
280 snSerTrpGlnGluTrpAsnIleTyrLysLysGluPheAlaAspGluIle 296
916 TACAGAGCGGATACACATACGCTCTNTTTGAACCGCGCAGTAACGACA 965
|||||
297 LysGlnArgAsp...AsnAlaGlyThrIleLysGlyTyrGlyGluHisH 312
966 TTTTTCCTTACATCCACACACACGTCGCGGTACGTAACAGAACCA 1015
|||||
312 sTrpLysThrThrGlyThrAsnSerHisIleGlySerThr..... 325
1016 ACGAAAAGGTNTCCAATCCAAAGCTTAAAGTACAGACAGTCGCACTGTT 1065
|||||
326 .....AlaValArgLeu... 329
1066 GACGATCTTTGATGAAGTAAAGAACCACTTTACCGCGCAGGGG 1115
|||||
330 .....AlaGlyAs 332
1116 TGTTAATCAGTCACCGTCCCAAGTTAAACACGCGTGAACCTTTCTTTA 1165
|||||
332 n.....GluArgGlyAlaAsnAsnGlyGlnAsnValThrPhe. 344
1166 TCGATTACGCAACGGCAACTCATCTTATCAAAACAACATCAACCAAGGC 1215
|||||
345 .....GluAsnAsnGlyThrLeuValLeuAspGlnAsnIleAsnGlnGly 359
1216 GCGGGCGGTTGTTTGAAGTGATTTTACGCTCTCGCCTGAAACAA 1265
|||||
360 AlaGlyGlyLeuPhePheLysGlyAspTyrThrValLysGlyAlaAsnAs 376
1266 CGAA...ACGTGCAAGCGCGCGCTTATATCATGTAAGACAGTACCG 1312
|||||

```

376 nGlyleThrTrpLeuGlyAlaGlyIleAspValAlaAspGlyLysLysV 393
1313 TTACTTGAAGTAACGGCGTGGCAACACGACCGCTGTCCAAATCGGC 1362
393 alValTpeGlnValLysAsnProAsnGlyAspArgLeuAlaLysIleGly 409
1363 AAAGCGCGTGCACGTTCAAGCCAAAGGGAACCAACCAAGCTCGATCAG 1412
410 LysGlyThrLeuGluIleAsnGlyThrGlyValAsnGlnGlyLysLys 426
1413 CGTGGCGACGCTACAGTCTATTTGGATCAGCAGCGACAGATGAAGGCA 1462
426 sValGlyAspGlyThrValIleLeuAsnGlnGlnAlaAspAlaAspLysL 443
1463 AAAACAAGCCTTAGTGAATCGCTTGNCTACCGCGCAGGCGGTACCGTG 1512
443 yValGlnAlaPheSerGlnValGlyIleValSerGlyArgGlyThrLeu 459
1513 CAACCTGAATCGCGATATACGTTCAACCCCGACAACTCTATTTCCGGCTT 1562
460 ValLeuAsnSerProAspGlnIleAsnProAsnAsnLeuTyrPheGlyPh 476
1563 TCGCGCGGACGCTTGGATTTAAACGGGCATTCGCTTCGTTCACCGTA 1612
476 eArgGlyGlyArgLeuAspAlaAsnGlyAsnAspLeuThrPheGluHisI 493
1613 TTCAAAATACCGATCAAGGCGGATGTTGNCNATCATATGCCACACA 1662
493 LeArgAsnValAspGluGlyAlaArgIleValAsnHisAsnThrAspArg 509
1663 ACATCCACCGTTACCATTTACAGGAATGAAGTATTACACAA..... 1704
510 AlaSerThrIleThrLeuThrGlyLysSerLeuIleThrAlaProGlnAs 526
1704 1704
526 nLeuSerValTyrGluIleArgAsnAspTyrAspAspAspTyrTyrG 543
1705CCAGTGGTAAGAAATATCAATAGA 1728
543 lYtTyrSerTyrArgLysProIleProGlnGlyLysAspLeuTyrTyr 559
1729 CTTAATTAC..... 1737
560 LysAsnTyrArgTyrTyrAlaLeuLysSerGlyGlySerValAsnAlaPr 576
1737 1737
576 oMetProGluAsnGlyGlnThrGluAsnAsnAspTrpIleLeuMetGlyS 593
1738AGCAAGAAATTCCTACAAC..... 1758
593 erThrGlnGluGluAlaLysLysAsnAlaMetAsnHisLysAsnAsnGln 609
1759GGTTGGTTTGGCGAAGATACACCAAAAC 1790
610 ArgIleSerGlyPheSerGlyPheGlyGlyGluAsnGlyLysGlyYhi 626
1791 GAACGGCGGCTCAACCTTGTTCACCGCCCGCCAGAGACCGCACCC 1840
626 sAsnGlyAlaLeuAsnLeuAsnPheAsnGlyLysSerAlaGlnAsnArgP 643
1841 NGCTGCTTTCCGCGGAACAAATTTAAACGGCAAGATCACGCAAAACAAAC 1890
643 heLeuLeuThrGlyThrAsnLeuAsnGlyLysIleSerValThrGln 659
1891 GGCAAACTGTTTACGGCGACAGCGACCGCGCTACATCAATTT 1940
660 GlyAsnValLeuLeuSerGlyArgProThrProHisAlaArgAspPheVa 676
1941 AGGAACGGGTGTGCAAAATGAAGGT.....ATCCCCAAGGAGAA 1984
676 lAsnLysSerSerAlaGlnLysAspAlaHisPheSerLysAsnAsnGluV 693

1985 TCGTGTGGACAACACTGGATCCNACCAGCGCTTTAAAGCGGAAATTC 2034
693 alValPheGluAspAspTrpIleAsnArgThrPheLysAlaThrGluIle 709
2035 CATATT...CAGGCGGCGAGCGGTGTATTTCCCGCAATGTGCCAAGT 2081
710 ThrValAsnGlnSerAlaSerPheSerSerGlyArgAsnValSerAsnI 726
2082 GGAAGCGGATTCNCATTTTCAGCAATCAGCCCAAGCAGTCTTTTGTGTCG 2131
726 eThrAlaAsnIleThrAlaThrAspAsnAlaLysValAsnLeuGlyTyr. 742
2132 CACCGCATCAAAAGCCATACATCTGTACAGCTTCGGACTGGACNGGTCTG 2181
743LysAsnGlyAspGluValCysValArgSerAspTyrThrGlyTyr 757
2182 ACAATTGTCTCGAANAANCATTTACCGCAGTAAAGTATTGCTTCATT 2231
758 ValThrCys..... 760
2232 GACTAAGACGACNTNAGCGCANTGTNAGCNCTNNCCNATNACGNTNNTT 2281
760 760
2282 NAAANCTCNCNGGCGTGCNACACTNNAAGCAATCTTAGTCAAATGCG 2331
761AsnThrGlyAsnLeuSer..... 766
2332 GATACAGTTATACGTACGCCACAACGCCACCCAA...AACGGCAACCT 2378
767 ...AspLysAlaLeuAsnSerPheGlyAlaThrGlnIleAsnGlyAsnVa 782
2379 TAGCCTCGTGGCAATGCCCAAGCAACATTTAATCAAGCACATTAACG 2428
782 lAsnLeuAsnGlnAsnAlaLeuValLeuGlyLysAlaLeuTrpG 799
2429 GCAACNATCGGNTTCGGCAATGCTTCATTTAATCAAGCAACACGCC 2478
799 lYnIleGlnGlyGlnGlyAsnSerArgValSerLeuAsnGlnHis... 814
2479 GCACAAAACGGCAGTCTGACGCTTCCGACACAGCTTAAGGCAACGTAAG 2528
814 814
2529 CCATTCCGCACTCAACGGCAATGCTCCCTAGCCGATAAGGCAAGTATTC 2578
814 814
2579 ATTTGAAAACAGCGCTTTACCGGACAACTCAGCGCAGCAAGANACA 2628
814 814
2629 GCAATTACACTTAAAGACAGCAATGGACGCTGCCGTTCAGGCACGGAATT 2678
815SerLysTrpHisLeuThrGlyAspSerGlnVa 825
2679 AGCAATTTAAACCTTGACAACGCCACCATACACTCAATTCGCGCTATC 2728
825 lHisAsnLeuSerLeuAlaAspSerHisIleHisLeuAsnAla... 840
2729 GCCAGATGTCGAGCGCGCAACCGCAGNGTGTTCAGACACGCCCGCGC 2778
841 ..SerAspAlaGlnSerAla..... 846
2779 CGCGCTTCGCGCGCTTCCTATTATTCGTTACACCGCAACTTCGCTGAGA 2828
846 846
2829 ATCCCGTTTCAACACGCTGAGGTAACCGCAATTAACNGTCAAGGAA 2878
847 .AsnGlnTyrHisThrLeuLysIleAsn...HisLeuSerGlyAsnGlyH 862

```

2879 CATTCGGCTTTATGTCGGAACCTCTCGCTACCGAAGCGCAAAATTGAAG 2928
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
862 lPheHistyLeuThrHisLeuAlaGluAsnLeuGlyAspLysValLeu 878
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
2929 CTGCGGGAAGTTCGGAAGNACTTACACCTTGGCGGTGCAACAATACCGG 2978
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
879 ValLysGluSerAlaSerGlyHisTyrGlnLeuHisValGlnAspLysTh 895
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
2979 CAAGGAACCCCTAAGCTCGATCAATGACGGTAGTGGAAGGGAAGACA 3028
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
895 rGlyGluPro...AsnGlnGluGlyLeuAsnLeuPheAspAlaSerSrv 911
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
3029 ACAACCGCTGTCGGAACAACTTAAATTCACCCCTGCAAAACGACAGTC 3078
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
911 alArGAspArgSer...HisLeuSerValSerLeuAlaAsnHisHisVal 926
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
3079 GATCGCGCGCTGGCTTACCACTCATCGGAAAGACGGCGAGTTCGG 3128
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
927 AspLeuGlyAlaLeuArgTyrThrIleLysThrGluAsnGlyIleThrAr 943
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
3129 CCTGCATAATCCGCTCAAGAAACAAGAGCTTCCGCAAACTCGGCAAGG 3178
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
943 gLeuTyrAsnProTyrAlaGluAsnArgArgValLysProAlaProS 960
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
3179 CAGAAGCC.....AAAAACAGCGGCAAAAGACACGCGCAA 3216
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
960 erProAlaThrAsnThrAlaSerGlnAlaGlnThrAspSerAlaGln 975
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
seq_name: sp_bacteria:030573
seq_documentation_block:
ID O30573 PRELIMINARY; PRT; 992 AA.
AC O30573;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE IGAL PROTEASE (FRAGMENT).
GN IGA.
OS Neisseria meningitidis.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=487;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Z3910;
RX MEDLINE=98010345; PubMed=9350862;
RA Morelli G., Malorny B., Muller K., Seiler A., Wang J.F., del Valle J.,
RA Achtman M.;
RT "Clonal descent and microevolution of Neisseria meningitidis during 30
RT years of epidemic spread";
RL Mol. Microbiol. 25:1047-1064(1997).
DR EMBL; AF012205; AAC45788.1; -.
DR MEROPS; S06.001; -.
DR InterPro; IPR002195; Dihydroorotase.
DR InterPro; IPR000710; IGA_S6.
DR Pfam; PF02395; IGAL; 1.
DR PRINTS; PR00921; IGASERPTASE.
DR PROSITE; PS00482; DIHYDROOROTASE_1; UNKNOWN_1.
KW Protease.
FT NON_TER 1 1
FT NON_TER 992 992
SQ SEQUENCE 992 AA; 109255 MW; EF7CA563BC7D7ECB CRC64;

```

```

alignment_scores:
  Quality: 1284.00      Length: 1149
  Ratio: 1.994          Gaps: 31
  Percent Similarity: 56.049  Percent Identity: 30.287

```

```
alignment_block:
```

```
US-09-303-518d-651 x O30573 ..
```

```
Align seg 1/1 to: O30573 from: 1 to: 992
```

```

94 CTGCTGTTCCGCATCTTCTCCCAAGCTTGGCGGGGACACACTATTTCGG 143
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
1 LeuAlaTyrAlaLeuThrProTyrSerGluAlaAlaLeuValArgAspAs 17
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
144 CATCAACTACCAATACTATCGGACTTTCGCGAAATAAAGCAAGTTTG 193
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
17 pValAspTyrGlnIlePheArgAspPheAlaGluAsnLysGlyLysPheP 34
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
194 CAGTCGGGCGGAAAGATATTGAGGTTNACACAAAAGGGAGTTGGTC 243
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
34 heValGlyAlaThrAspLeuSerValLysAsnLysGlnGlyGlnAsnIle 50
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
244 GCAATATCAATGACAAAACCCCGATGATTGATTTTCTGTGTCGTCGG 293
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
51 GlyAsnAlaLeuSerAsnValProMetIleAspPheSerValAlaAspVa 67
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
294 TAAC...GGCGTGGGCGCATTTGGTGGCGGATCAATATATTGTGACGTCG 340
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
67 lAsnArgArgThrLeuThrValIleAspProGlnTyrAlaValServalL 84
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
341 CACATAACGGCGGTATACAAACGTTGATTTTGGTCGGAAGAAAGNAAT 390
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
84 yHisValLysGlyAspGluIleSerTyrTyrGlyHisHisAsnGlyHis 100
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
391 CCGGAT.....CAGCACCGTTTCTTACCAAAATTTGAAAGAAATAA 434
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
101 LeuAspValSerAsnAspGluAsnGluTyrArgSerValAlaGlnAsnAs 117
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
435 TTATAAGCTGACAAATTCACCTTTACACACGCG.....G 469
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
117 pTyrGluProAsnLysAsnTrpHisHisGlyAsnGlnGlyArgLeuGlu 134
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
470 ATTANCATATCGCGGTTTGCATAAATTTGTCACAGATGCAGAACCTGTC 519
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
134 sPtyrAsnMetAlaArgLeuAsnLysPheValThrGluValAlaProIle 150
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
520 GAAATGACGAGTGACATGAGGGGG...AATACCTATTCCGATAAGAA 566
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
151 AlaProThrSerAlaGlyGlyValGluThrTyrLysAspLysAsnAr 167
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
567 ATATCCCGCGGTGTCGCATCGGTCAGGACACCACTATTTCGGGTTATG 616
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
167 gPheSerGluPheValArgValGlyAlaGlyThrGlnPheGluTyrAsnS 184
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
617 ATGATGACAAACACGCGCATTTATCCCTACTCCGGCGCATGTTTATGCG 666
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
184 erArgTyrAsnMetThrGluLeuSerArgAlaTyrArgTyrAlaIleAla 200
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
667 GCAATACACATATGACG..... 684
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
201 GlyThrProTyrGlnAspValAsnValThrSerAsnLeuAsnGlnGluG 217
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
685 .....GTTGGGGAATAATGCGGTANTTAGTTTACGCGCGCATGTCG 727
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
217 yLeuIleGlyPheGlyAspAsnSerLysHisHisSerLeuGluLysLeu 234
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
728 GCCATGCCAAGCACTATGCGCCCTATGCCGATTCGAGGTGCGCGCAGG 777
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
234 ysGluValLeuSerGlnAsnAlaLeuThrAsnTyrAlaValLeuGlyAsp 250
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
778 AGCGGTTCCCAATGTTTATTATGACAAAACAAATAAATGCGTGCT 827
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
251 SerGlySerProLeuPheAlaTyrAspLysGlnGlyLysArgTyrPValPh 267
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
828 CAACGGAGTTTACAA.....ACCGGTACCTCTTATTCGGGCGAGGAAA 871
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
267 eLeuGlyAlaTyrAspTyrTrpAlaGlyTyr.....GlnLysA 280
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
872 ACGGTTTCCAG.....CTGATACGCAAAAGATTGGTTCTACGATGACATT 915
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
280 snSerTrpGlnCluTrpAsnIleTyrLysLysGluPheAlaAspGluIle 296
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
916 TACAGAGGCGATACACATACGCTCTNTTTTGAACCGCGCAGTAACGGACA 965
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

```

297 LysGlnArgAsp...AsnAlaGlyThrIleLysGlyTyrGlyGluHisI 312
 966 TTTTTCCTTACATCCACAACAACGGGTACGGGTACGGTACGTAACAACAACCA 1015
 312 strLysThrThrGlyThrAsnSerHisIleGlySerThr... 325
 1016 ACGAAAAGGTNTCCAAATCCAAAGCTTAAAGTACAGACAGTCCGACTGTTT 1065
 326AlaValArgLeu... 329
 1066 GACGAATCTTTGAATGAACACTGATAAAGACAGTTTACGGCGGAGGGG 1115
 330AlaGlyAs 332
 1116 TGTTAATCAGTACCGTCCAAAGGTAAACAACGGTGAACAACTTCTTTT 1165
 332 n.....GluArgGlyAlaAsnAsnGlyGlnAsnValThrPhe. 344
 11166 TCGATTACGGCAACGGCAAACTCATCTTATCAACAACATCAACCAAGGC 1215
 345GluAsnAsnGlyThrLeuValLeuAspGlnAsnIleAsnGlnGly 359
 1216 GCGGCGGTTTGTATTGAGGTGATTTACGGTCTCGCCTGAAACAA 1265
 360 AlaGlyGlyLeuPhePheLysGlyAspTyrThrValLysGlyAlaAsnAs 376
 1266 CGAA...ACGTGGCAAGCGCGCGCTTCATCAGTGAAGACAGTACCG 1312
 376 nGlyIleThrTrpLeuGlyAlaGlyIleAspValAlaAspGlyLysLysV 393
 1313 TTACTTGGAAAGTAACGGCGTGGCAACAGCCGCTGTCCAAATCCGC 1362
 393 alValTrpGlnValLysAsnProAsnGlyAspArgLeuAlaLysIleGly 409
 1363 AAAGCGACGTGACGTTCAAGCAAGGGGAAACCAAGGTCGATCAG 1412
 410 LysGlyThrLeuGluIleAsnGlyThrGlyValAsnGlnGlyGlnLeuLy 426
 1413 CGTGGCGACGGTACAGTCAATTTGGATCAGCAGGCAAGTAAAGCA 1462
 426 sValGlyAspGlyThrValIleLeuAsnGlnAlaAspAlaAspLysL 443
 1463 AAAACAAGCCTTAGTGAATCGGCTTGTACAGCGGCAAGGTCAGGTG 1512
 443 ysValGlnAlaPheSerGlnValGlyIleValSerGlyArgGlyThrLeu 459
 1513 CAACTGAATCCGATATCAGTTCAACCCCGCAACTCTATTTCGGCTT 1562
 460 ValLeuAsnSerProAspGlnIleAsnProAsnAsnLeuTyrPheGlyPh 476
 1563 TCGCGGCGGAGTTGATTTAAACGGGATTCGCTTTCGTTCCACCGTA 1612
 476 eArgGlyGlyArgLeuAspAlaAsnGlyAsnAspLeuThrPheGluHisI 493
 1613 TTCAAAATACGATGAAGCGCGGATGATTCGNCNATCATATAATGCCAACA 1662
 493 leArgAsnValAspGluGlyAlaArgIleValAsnHisAsnThrAspArg 509
 1663 ACATCCACCGTTACCATCAGCGAATGAAGTATTACAA... 1704
 510 AlaSerThrIleThrLeuThrGlyLysSerLeuIleThrAlaProGlnAs 526
 1704 1704
 526 nLeuSerValTyrGluIleArgAsnAspTyrAspAspAspTyrTyrg 543
 1705CCGAGTGGTAAGATATCAATAGA 1728
 543 lyTyrTyrSerTyrArgLysProIleProGlnGlyLysLeuTyrTy 559
 1729 CTTAATTAC... 1737

```

2529 CCATTCCGCACTCAACGGCAATGCTCTCCCTAGCCGATAGGCAAGTATTC 2578
814 ..... 814
2579 ATTTTGAACAACAGCGGCTTTACCGGACAACCTACGCGGCAAGGANACA 2628
814 ..... 814
2629 GCATTACACTTAAAGACACGGAATGGACGCTGCCGTCAGGCACGAAAT 2678
815 ..... 815
2679 AGCAATTTAAACCTTGACACACGCAATTCACACTCAATTCGCGCTATC 2728
825 lHisAsnLeuSerLeuAlaAspSerHisIleHisLeuAsnAla.... 840
2729 GCACAGATGTCGAGCGCGCAACACGCGCAGNGTGTGACACACGCGCGC 2778
841 ..SerAspAlaGlnSerAla..... 846
2779 CGCGTTTCGCGCGTTCCTTATTCGTTACACGCCCACTTCGCTAGA 2828
846 ..... 846
2829 ATCCGTTTCAACACGCTGACGGAACGCAAAATGCAACNGTCAAGAA 2878
847 .AsnGlnTyrHisThrLeuLysIleAsn...HisLeuSerGlyAsnGlyH 862
2879 CATTCGCTTATGTCGGAACCTCTCGGTCACGGAACGCAAAATGGAAG 2928
862 isPheHisTyrLeuThrHisLeuAlaGluAsnLeuGlyAspLysValLeu 878
2929 CTGCGGAAAGTTCGGAAGNACTTACACCTTCGCGGTCAACATACCGG 2978
879 ValLysGluSerAlaSerGlyHisTyrGlnLeuHisValGlnAspLys 895
2979 CAACGAACCGGTAAAGCTCAATGAGCGTAGTGAAGGAAAGACA 3028
895 rGlyGluPro...AsnGlnGluGlyLeuAsnLeuPheAspAlaSerV 911
3029 ACAACCGCTGTCGGAACCTTAATTCACCTGCAAAACGCAACACGTC 3078
911 alArgAspArgSer...HisLeuSerValSerLeuAlaAsnHisIleVal 926
3079 GATCGCGCGCGTGGGCTTACCAACTCATCCGCAAAAGACGCGAGTCCG 3128
927 AspLeuGlyAlaLeuArgTyrThrIleLysThrGluAsnGlyIleThrAr 943
3129 CTGCGATAATCCGGTCAAGACAAGAGCTTTCGCAAACTCGGCAAGG 3178
943 gLeuTyrAsnProTyrAlaGluAsnArgArgValLysProAlaProS 960
3179 CAGAACCC.....AAAAACAGCGGGAACACACACGCGCAA 3216
960 erProAlaThrAsnThrAlaSerGlnAlaGlnThrAspSerAlaGln 975

```

seq_name: sp_bacteria:Q956X5

seq_documentation_block:

ID_Q956X5 PRELIMINARY; PRT; 992 AA.

AC Q956X5;

DT 01-MAY-2000 (TREMBLrel. 13, Created)

DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)

DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)

DE IGA1 PROTEASE (FRAGMENT).

GN IGA.

OS Neisseria meningitidis.

OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.

OX NCBI_TaxID=487;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Z2491;

```

RX MEDLINE=98010345; PubMed=9350862;
RA Morelli G., Malorny B., Muller K., Seiler A., Wang J.F., del Valle J.,
RT Achtmann M.;
RT "Clonal descent and microevolution of Neisseria meningitidis during 30
RL years of epidemic spread.";
RL Mol. Microbiol. 25:1047-1064(1997).
DR EMBL: AF012203; AAC45786.1;
DR MEROPS: S06.001;
DR InterPro: IPR002195; Dihydroorotase.
DR InterPro: IPR000710; Iga_S6.
DR Pfam: PF02395; IGA1; 1.
DR PRINTS: PS00921; IGASERPTASE.
DR PROSITE: PS00482; DIHYDROOROTASE_1; UNKNOWN_1.
KW Protease.
FT NON_TER 1
FT NON_TER 992
SQ SEQUENCE 992 AA; 109239 MW; 306FF451B7213D49 CRC64;

```

alignment_scores:

Quality: 1281.00 Length: 1149
 Ratio: 1.989 Gaps: 31
 Percent Similarity: 56.049 Percent Identity: 30.200

alignment_block:

US-09-303-518D-651 x Q956X5

Align seg 1/1 to: Q956X5 from: 1 to: 992

```

94 CTGCTGCTCGCATTCCTCCCAAGCTTGGCGGACACACTTATTCGG 143
1 LeuAlaTyrAlaLeuThrProTyrSerGluAlaAlaLeuValArgAspAs 17
144 CATCAACTACCACTACTATCGGACTTTGCCGAAATAAAGCAAGTTTG 193
17 pValAspTyrGlnIlePheArgAspPheAlaGluAsnLysGlyLysPheP 34
194 CAGTCGGCGGCAAGATATTTGAGGNTACACAAAAAGGGAGTTGTC 243
34 heValGlyAlaThrAspLeuSerValLysAsnLysGlnGlyClnAsnIle 50
244 GCAATCAATGACAAAAAGCCCGATGATTCTTCTGTGGTGTCCGCG 293
51 GlyAsnAlaLeuSerAsnValProMetIleAspPheSerValAlaAspVa 67
294 TAAC...GGCTGGCGGCGATTCGGTGGCGATCAATATATTGTGACCGTGG 340
67 IasnArgArgThrLeuThrValIleAspProGlnTyrAlaValSerVal 84
341 CACATAACGGCGCTATAACCAACGTTGATTTTGGTGGCAAGAGNAAT 390
84 yHisValLysGlyAspGluIleSerTyrTyrGlyHisHisAsnGlyHis 100
391 CCCGAT.....CAGCACCGCTTTTCTTACCATAATTTGTGAAAAGATAA 434
101 LeuAspValSerAsnAspGluAsnGlnTyrArgSerValAlaGlnAsnAs 117
435 TTATAAGCCTGACAATTC.....CACCCCTTACAACGCG.....G 469
117 pTyrGluProAsnLysAsnTrpHisHisGlyAsnGlnGlyArgLeuGluA 134
470 ATTANCATATCGCGCTTTCATATAATTTGTACAGATGTCAGAACCTGTC 519
134 spTyrAsnMetAlaArgLeuAsnLysPheValThrGluValAlaProIle 150
520 GAAATGACGAGTGCATGAGGGG...AATACCTATTCCGATAAAGAAA 566
151 AlaProThrSerAlaGlyGlyValGlnThrTyrLysAspLysAsnAr 167
567 ATATCCCGAGCGTTCGCGATCGCTCAGGACACCACTATTGCGGTATG 616
167 gPheSerGluPheValArgValGlyAlaGlyThrGlnPheGluTyrAsnS 184

```



```
743 .....LysAsnGlyAspGluValCysValArgSerAspTyrThrGlyTyr 757
2182 ACAATTGTGTGCGAANAANCAATTACCGACGATAAAAGTATTGCTTCATT 2231
       : : : : :
758 ValThrCys..... 760
2232 GACTAAGACGACGACNTNAGCGGCANTGTNAGNCTNNCCNATNAGGNTNNTT 2281
760 ..... 760
2282 NAAANCTCNCNGGCGTGCNNCACTNAGNGCAATCTTAGTCAAAATGGC 2331
761 .....AsnThrGlyAsnLeuSer..... 766
2332 GATACAGTTATACAGTCAGCCACAAGCCACCACAA...AACGGCAACCT 2378
767 ...AspLysAlaLeuAsnSerPheGlyAlaThrLysIleAsnGlyAsnVa 782
2379 TAGCGTCGTGGGCAATGCCCAAGCAACATTTAATCAAGCCACATTAACG 2428
782 IAsnLeuAsnGlnAsnAlaLeuValLeuGlyLysAlaLeuTyrG 799
2429 GCAACNATCGGNTTCGGCAATGCTTCATTTAATCTAAGCAACACGCGC 2478
799 LysGlnIleGlnGlyGlnGlyAsnSerArgValSerLeuAsnGlnHis... 814
2479 GCACAAACGGCAGTCTGACGCTTCCGACACGCTAAGGCAACGTAAG 2528
814 ..... 814
2529 CCATTCGGCACTCAACGGCAATCTCCCTAGCCGATAAGCGAGTATCC 2578
814 ..... 814
2579 ATTTGAAAACAGCGCTTTACCGGACAACTCAGCGCAGCAAGGANAACA 2628
814 ..... 814
2629 GCATTACACTTAAAGACAGCAATGAGCGCTGCCGTACGACGCGAAT 2678
815 .....SerLysTyrHisLeuThrGlyAspSerGlnVa 825
2679 AGCAATTTAAACCTTGACACGCCACCAATTACACTCAATTCGGCTATC 2728
825 IHisAsnLeuSerLeuAlaAspSerHisIleHisLeuAsnAsnAla... 840
2729 GCCACGATGCTGACGCGCGCAACCGCGAGNGTGTGCACACACGCGCGC 2778
841 ...SerAspAlaGlnSerAla..... 846
2779 CGCGGTTCGCGCGGTTCCTATTATCCGTTACACGCCCAACTTCGGTAGA 2828
846 ..... 846
2829 ATCCCGTTTCAACAGCTCAGCGTAAACGGCAAAATTTGAACNCTCAAGGAA 2878
847 .AsnGlnTyrHisThrLeuLysIleAsn...HisLeuSerGlyAsnGlyH 862
2879 CATTCGCTTTATGTCGGAACTCTTCGGCTACCGAAGCGACAAATTTGAAG 2928
862 isPheHisTyrLeuThrHisLeuAlaGluAsnLeuGlyAspLysValLeu 878
2929 CTGGCGGAAGTTCCGAGNACTTACACCTTGGCGGTCAACAAATACCGG 2978
879 ValLysGluSerAlaSerGlyHisTyrGlnLeuHisValGlnAspLysTh 895
2979 CAACGAACCCGTAAGCTTCGATCAATTTACGGTACTGGAAGGGAAGACA 3028
895 rGlyGluPro...AsnGlnGluGlyLeuAsnLeuPheAspAlaSerSerV 911
3029 ACAAAACCGCTGTCCGAAACCTTAATTTACCCCTGCAAAACGACACGCTC 3078
911 aIArgAspArgSer...HisLeuSerValSerLeuAlaAsnHisHisVal 926
```

```
3079 GATGCCGGCGCGTTCGCTTACCAACTCATCCCAAGACGGCGAGTTCG 3128
       : : : : :
927 AspLeuGlyAlaLeuArgTyrThrIleLysThrGluAsnGlyIleThrAr 943
3129 CTGTGATATCCGTCARAAGAACAGCTTTCGACAACTCGCAAGG 3178
       : : : : :
943 gluLysAsnProTyrAlaGluAsnArgArgValLysProAlaProS 960
3179 CAGAAGCC.....AAAAACAGGGGAAAAAGACAAACGCGCAA 3216
       : : : : :
960 exProAlaThrAsnThrAlaSerGlnAlaGlnThrAspSerAlaGln 975

seq_name: sp_bacteria:Q47692

seq_documentation_block:
ID Q47692 PRELIMINARY; PRT: 1377 AA.
AC Q47692;
DT 01-NOV-1996 (T-EMBLrel. 01, Created)
DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)
DE TSH.
GN Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CHI7122;
RX MEDLINE=94178945; PubMed=8132344;
RA Provence D.L., Curtiss R. III;
RT "Isolation and characterization of a gene involved in hemagglutination
RL by an avian pathogenic Escherichia coli strain.";
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CHI7122;
RX MEDLINE=20316042; PubMed=10858231;
RA Dozois C.M., Dho-Moulin M., Bree A., Fairbrother J.M., Desautels C.,
RA Curtiss R. III;
RT "Relationship between the Tsh autotransporter and pathogenicity of
RT avian Escherichia coli and localization and analysis of the Tsh
RT genetic region.";
RL Infect. Immun. 68:4145-4154(2000).
DR EMBL; AF218073; AAA24698.1; -
DR MEROPS; S06.003; -
DR InterPro; IPR000710; IGA_S6.
DR Pfam; PF02395; IGAL; 1.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
KW Hydrolase; Plasmid; Serine protease.
SQ SEQUENCE 1377 AA; 148226 MW; 237423644D9AE012 CRC64;

alignment_scores:
Quality: 1084.50 Length: 1508
Ratio: 1.378 Gaps: 54
Percent Similarity: 52.188 Percent Identity: 25.995

alignment_block:
US-09-303-518D-651 x Q47692
Align seg 1/1 to: Q47692 from: 1 to: 1377

90 ATGCTCTGCTTCGGCAATTCCTCCCAAGCTTGGCGGGACACACTTATT 139
   : : : : :
42 IleProValLeu...PheSerAlaGlySerLeuAlaGlyThr.ValAsnA 57
140 TCGCATCACTACCAATACATATCGGACATTCGCCAAAATAAGGCAAG 189
   : : : : :
57 snGluLeuGlyTyrGlnLeuPheArgAspPheAlaGluAsnLysGlyMet 73
```

1066 GACGAATCTTTGAATGAATACTGATAAAGAACAGTTCAGCGCGGACGGGG 1115
340ThrThrTyrAlaMetHisG1 346
1116 TGTATATAGTACCTCCAGCTTAACAACGCTGAACACCTTCTCTTTA 1165
346 y.....GlnGlnGlyAsnAspLeuAsnAlaGlyLysAsnLeuIlePhe. 360
1166 TCGATTACGGCAACGCAACTCATCTTATCAACAACATCAACCAAGCG 1215
361 ..GlnGlyGlnAsnGlyGlnIleAsnLeuLysAspSerValSerGlnGly 376
1216 GCGGCGGTTTGTATTTGAAGGTGATTTACGGTCTCCCTGAAACAA 1265
377 AlaGlySerLeuThrPheArgAspAspTyrThrValThrSerAsnG1 393
1266 CGAAACGTGCGAAGCGCGGCTTCATATCAGTGAAGACAGTACCGTTA 1315
393 ySerThrThrThrGlyAlaGlyIleValValAspAsnGlyValSerVala 410
1316 CTTGGAAGTAACGCGCTGCAACAGCGCTGTCCAAATTCGGCAAA 1365
410 snTrpGlnValAsnGlyValLysGlyAspAsnLeuHisLysIleGlyGlu 426
1366 GGCACGCTGCAGCTTCAAGCCAAAGGGGAAACCAAGCGCTCGATCAGCGT 1415
427 GlyThrLeuThrValGlnGlyThrGlyIleAsnGluGlyGlyLeuLysVa 443
1416 GCGCGAGGTACAGTCACTTTTGGATCAGCAGGCGACAGCAGTAAAGGCAAAA 1465
443 lGlyAspGlyLysValValLeuAsnGlnGlnAlaAspAsnLysGlyGlnV 460
1466 AACACGCTTTAGTGAAATCGGCTTNTCAGCGCGGAGGTACGGTGCAA 1515
460 alGlnAlaPheSerSerValAsnIleAlaSerGlyArgProThrValVal 476
1516 CTGAATCCCGATTAATCAGTTCACCCGCAAACTCTATTTCGGCTTCG 1565
477 LeuThrAspGluArgGlnValAsnProAspThrValSerTrpGlyTyrAr 493
1566 CCGCGGACGTTTGGATTAAACGCGCATTCGGTTTCCTCCACCGATTTC 1615
493 gGlyGlyThrLeuAspValAsnGlyAsnSerLeuThrPheHisGlnLeuL 510
1616 AAATACCGATGAGGGCGATGATTCNCNATCATTAATGCCACAACAACA 1665
510 ysaAlaAlaAspTyrGlyAlaValLeu...AlaAsnAsnValAspLysArg 525
1666 TCCACCGTTACCATTCAGGGAATGAAGTATTACACACCGAGTGTAA 1715
526 AlaThrIleThr..... 529
1716 GAATATCAATAGACTTAATTAC.....AGCAAGAATAATTCCTACA 1756
530LeuAspTyrAlaLeuArgAlaAspLysValAlaLeuA 542
1757 ACGGTTGG..... 1764
542 snGlyTrpSerGluSerGlyGlyThrAlaGlyAsnLeuTyrLysTyr 558
1765TTTGGCGAGAAAGATACGAC 1784
559 AsnAsnProTyrThrAsnThrThrAspTyrPheIleLeuLysGlnSerTh 575
1785 C..... 1785
575 rTyrGlyTyrPheProThrAspGlnSerSerAsnAlaThrTrpGluPheV 592
1785 1785
592 alGlyHisSerGlnGlyAspAlaGlnLysLeuValAlaAspArgPheAsn 608
1786AAAACGAACGCGCGCTCAACCTGT 1811

190 TTTGAGTCGCGGCAAGATATTGAGTNTACAAACAAAAAGGGAGTT 239
74 PheArgProGlyAlaThrAsnIleAlaIleTyrAsnLysGlnGlyGluPh 90
240 GGTGGCAAAATCAATGACAAAAGCCGATGATTGATTTTCTGTGTG. 288
90 eValGly...ThrLeuAspLysAlaAlaMetProAspPheSerAlaVala 106
289 ..TCGCGTAACGGGTGCGGCATTTGGTGGCGATCAATATATTGTGAGC 336
106 spSerGluIleGlyValAlaThrLeuIleAsnProGlnTyrIleAlaSer 122
337 GTGGCACATAACGGCGGTATACAAACGCTGATTTTGTGCGGAAGGAAG 386
123 ValLysHisAsnGlyGlyTyrThrAsnValSerPheGly...AspGlyG1 138
387 NAATCCCGATCAGCACCGCTTTTCTTACCAAAATGTGAAAGAATAATT 436
138 uAsn.....ArgTyrAsnIleValAspArgAsnAsnA 149
437 ATAAGCCTGCAATTCACACCTTACACGCGGATTCATATATCCCGGT 486
149 laPro.....SerLeuAspPheHisAlaProArg 158
487 TTGCATAAATTTGCACAGATCGAAGCTGTGCAAAATGACAGTGAC.. 534
159 LeuAspLysLeuValThrGluValAlaProThrAlaValThrAlaGlnG1 175
535 ..ATGAGGGGAATACCTATTCGGTAAGAAGAAATATCCCGAGGTGCC 583
175 yAlaValAlaGlyAlaTyrLeuAspLysGluArgTyrProValPheTyrA 192
584 GCATCGGTCGAGCACACCTATTGGCGCTTATGATGATGACAAACACGCG 633
192 rGluGlySerGlyThrGlnTyrIleLys.....AspSerAsnGly 205
634 GATTATCTACTCCGCGC.....GCATGTTTAATTTGGCGCAATAC 674
206 GlnLeuThrGlnMetGlyGlyAlaTyrSerTrpLeuThrGlyGlyThrVa 222
675 ACATATCGAGGTGGGGAATAATGCG.....GTANTTACTTTGACGG 718
222 lGlySerLeuSerTyrGlnAsnGlyGluMetIleSerThrSerSerG 239
719 GCGATGTGCGCCATGCCAACGACTATGCGCTATCCGATTCGAGGTGCG 768
239 lYLeuValPheAspTyrLysLeuAsnGlyAlaMetProIleTyrGlyGlu 255
769 GCAGGCGAGCGGTTCGCCAATGTTTATTATGACAAACAAACAATAA 818
256 AlaGlyAspSerGlySerProLeuPheAlaPheAspThrValGlnAsnLy 272
819 ATGGCTGCTCAACGGAGTTTACAAACGCGTACCTTATTCCGCGCAGG 868
272 strpValLeuValGlyValLeuThrAlaGlyAsnGlyAlaGlyGlyArgG 289
869 AAAACGTTTCCAGCTGATACCAAGATTTGTTCTACGATGACATTTAC 918
289 lYAsnAsnTrpAlaValIleProLeuAspPheIleGlyGlnLysPheAsn 305
919 AGAGGCGATACATACCTGCTNTTTGNAACCGCGCAGTAACGACATTT 968
306 GluAspAsnAlaProValThrPheArgThrSerGluGlyGlyAlaLe 322
969 TTCCTTTTACATCAACAACAAC...GGTACGGGTACGGTAAACGAACA 1015
322 uGluTrpSerPheAsnSerSerThrGlyAlaGlyAlaLeuThrGlnGlyT 339
1016 ACGAAAAGGTNTCCAAATCCAAAGCTTAAAGTACAGACAGTCCGACTGTT 1065
339 hr..... 339

609 ThrAlaGlyTyrLeuPheHisGlyGlnLeuLysGlyAsnLeuAsnValas 625

1812 TTACCAG...CCGCGCGAGACCGCAGCCAGCCTGCTTTCGCGCGAA 1858

625 pAsnArgLeuProGluGlyValThrGlyAlaLeuValMetAspGlyAlaA 642

1859 CAAATTAAACGGCAACATCAGCAACAAACGCAACCTGTTTTCAGC 1908

642 laAspIleSerGlyThrPheThrGlnGluAsnGlyArgLeuThrLeuGln 658

1909 GGCAGACCGACCGACCGACCGCTACAAATCAT 1938

659 GlyHisProValIleHisAlaTyrAsnThrGlnSerValAlaAspLysLe 675

1939 TTAGGAACGGGTGGTCAAAANTGGAAGGTATCCACAGGAGAAATCG 1987

675 uAlaAlaSerGlyAspHisSerVal.....LeuThrGlnProThrSerP 690

1988 TGTGGGACGACGATCGATCNACCGCAGCTTTAAAGCGGAAAAATTTCAT 2037

690 heSerGlnGluAspTrpGluAsnArgSerPheThrPheAspArgLeuSer 706

2038 ATTACGGCGCGCAGCGCGTGTATTCGCGCAATGTT.....GCCAA 2078

707 LeuLysAsnThrAspPheGlyLeuGlyArgAsnAlaThrLeuAsnThrTh 723

2079 AGTGAAGCGGATTCNCATTTTGAGCAATCAGCCCAAGCAGCTTTTGGTG 2128

723 rIleGlnAlaAsp.....AsnSerSerValThrLeuGly. 734

2129 TCGCACCGCATCAAAAGCCATACATCTGTACACCTTCGGACTCGAANGGT 2178

735AspSerArgValPheIleAspLysAsnAspGlyGlnGly 747

2179 CTGACAAATTTGTCTCGAA.....NAAACATTTACCGACGATAAAGT 2219

748 ThrAlaPheThrLeuGluGluGlyThrSerValAlaThrLysAsp..... 762

2220 GATTGCTTCATTGACTAAGACGACGANTNAGCGGCANTGTNAGCTNNCCN 2269

763AlaAspLysSerValPheAsnGlyThrValAsnLeuAspA 776

2270 ATNACGNTNNTNAAACNTCNCGGGGNTGNCNCACTNAAAGCACAATCTT 2319

776 snGlnSerValLeuAsnIleAsnAsp.....IlePheAsnGlyGlyIle 790

2320 AGTGCAAATTTGGCGATACAGCTTATACAGTCAGCACACACCCACCCAAA 2369

791 GlnAlaAsnAsnSerThrValAsnIleSerSerAspSerAla..... 804

2370 CGGCAACCTTAGCTCGTGGGCAATGCCCAAGCAACATTTTAATCAAGCCA 2419

805ValLeuGlyAsnSerThrLeuThrSerThrAlaLeuA 817

2420 CATTTAAACGGCAACATCGNNTTCGGCAATGCTTCATTT..... 2460

817 snLeuAsnLysGlyAlaAsnAlaLeuAlaSerGlnSerPheValSerAsp 833

2461AATCTAAGCAACACCGCCGCCAA..... 2484

834 GlyProValAsnIleSerAspAlaAlaLeuSerLeuAsnSerArgProAs 850

2485ACGGCAGTCTGA 2497

850 pGluValSerHisThrLeuLeuProValTyrAspTyrAlaGlySerTrpA 867

2498 CGCTT...TCGCACAAACGCTAAGCAACACGTAAGC...CATTCGCGACTC 2541

867 snLeuLysGlyAspAspAlaArgLeuAsnValGlyProTyrSerMetLeu 883

2542 AACGGCAATGCTCCCTAGCGGATAAG..... 2568

884 SerGlyAsnIleAsnValGlnAspLysGlyThrValThrLeuGlyGlyG1 900
2569
900 uGlyGluLeuSerProAspLeuThrLeuGlnAsnGlnMetLeuTyrSerL 917
2573 TATTCCATTTTGAACACAGCCGCTTTACCGGCAAACTCAGCGGACGAAG 2622
:||||:
917 eupHeAsnGlyTyArgAsnIleTyrSerGlySerLeuAsnAlaProAsp 933
2623 GANACACCATACACTTAAAGACACAGCGAATGAGCGCTGCCTCAGCAC 2672
:||||:
934 AlaThr...ValSerMetThrAspThrGlnTyrSerMetAsnGlyAsnSe 949
2673 GGAATTAGGCAATTAAACCTTTGACACGCGCACCATTACACTCAATTCG 2722
|||||:
949 rThrAlaGlyAsnMetLysLeuAsnArgThrIleValGlyPheAsnGly 966
2723 CWTATCGCCACGATGTCGACGCGCGCAAAACCGCAGNGTGTCAAGACG 2772
:|
966 Iy..... 966
2773 CCGGCGCGGGTTCGGCGCGTTCCTATATTCGGTTACAGCGCCAACTTC 2822
966 966
2823 GGTAGAATCCCGTTTCAACACGCTCAGCGTAACGCGCAATTTGAACNGTC 2872
||| |||||:
967ThrSerProPheThrLeuThrThrAspAsnLeuAspAlaValG 982
2873 AAGGAACATTCGCGCTTTATGTGCGAACTCTTCGGCTACCGAAGCGACAAA 2922
|||: |||
982 InSerAlaPheValMetArgThrAspLeu.....AsnLysAlaAspLys 996
2923 TTCAAGCTGGGGAAGTTCGGAAGNACTTACACCTTGGCGGTGCAACAA 2972
LeuValIleAsnLysSerAlaThrGlyHisAspAsnSerIleTyrPvalAs 1013
2973 TACCGGCAACGAAACCGTAAGCCTCGATCAATTG 3006
|
1013 nPheLeuLysLysProSerAsnLysAspThrLeuAspIleProLeuValS 1030
3007ACGGTA 3012
1030 erAlaProGluAlaThrAlaAspAsnLeuPheArgAlaSerThrArgVal 1046
3013 GTGGGAAGGAAGACACAAACCGGTGTCGAAACCTTAAATTCACCCCT 3062
|||: |||
1047 ValGlyPheSerAspValThrProfileLeuSerValArgLysGluAspG1 1063
3063 GCAAAACGAACAGTCGATGCGCGCGTGCCTTACCAACTCATCCGCA 3112
:||||: |||
1063 yLysLysGluTyrPvalLeuAspGly.....TyrGlnValAlaArgA 1077
3113 AAGACGGCGAGTTCGCGCTGCATATATCGGTCAAGAACAAAGAGCTTCC 3162
:||||:
1077 snAspGlyGln..... 1080
3163 GACAACTCGCAAGCGACAGACCAAAACACAGCGGAAAGACAAAC... 3210
|||||: |||
1081GlyLysAlaAlaThrPheMetHisIleSerTyrAsnAs 1094
3211GGCAAAACGCTTGACGCGCTGATTGCGCGCGCGCGG 3247
:||||: |||
1094 nPheIleThrGluValAsnAsnLeuAsnLysArgMetGlyAspLeuArgA 1111
3248 ATGCCGCCGAAAGACAGAAAGCTTCCGAACCGCGCCGCCGNGCAGGC 3297
||
1111 spIleAsnGlyGluAlaGlyThrTrpValArgLeuLeuAsnGlySerGly 1127
3298 GGGGAAATGTCGC.....ATTATGACGCGCGGAGGA 3329
:|: |||
1128 SerAlaAspGlyGlyPheThrAspHisTyrThrLeuLeuGlnMetGlyAl 1144

[illegible]

285 nglyAlaThrAsnTrpAsnValIleProThrAsp...TyrLeuAsnG 301
911 ACATTTACAGAGGATACACATACACCTCTNTTTTGAACCGCAGTAAC 960
301 InValMetGlnAspAspPheAspAlaProValAspPheValSerGlyLeu 317
961 GGACATTTTCTTTACA...TCCAAACAACAGGTACGGTACGGTAAC 1007
318 GlyProLeuAsnTrpThrTyrAspLysThrSerGlyThrGlyThrLeuSe 334
1008 AGAACCAACCAAGAGTNTCCATCCAAAGCTTAAAGTACACAGCTCC 1057
334 rGlnGlySer...LysAsnTrpThrMeth 343
1058 CACTGTTTGACGAATCTTTGAATGAACACTGATAAAGAACACCTTTACGCG 1107
343 is.....GlyGlnLysAspAsnAsp..... 349
1108 GCAGGGGTGTTAATCAGTACCGTCCAAAGTTAAACAACGGTGAACCT 1157
350LeuAsnAlaGlyGluAsnLe 356
1158 TCTTTTATCATTACGCAACGCAACCTCATCTTATCAACAACATCA 1207
356 uValPheSerGlyGln...AsnGlyAlaIleIleLeuLysAspSerValT 372
1208 ACCAAGCGCGGGGTTGTTATTTGAAGTGATTTACGGTCTCGCT 1257
372 hrGlnGlyAlaGlyTyrLeuGluPheLysAspSerTyrProValSerAla 388
1258 GAAACAACGAACGTCGCAAGCGCGCGCTTCATATCAGTGAAGACAG 1307
389 GluSerGlyLysThrTrpThrGlyAlaGlyIleIleThrAspLysGlyTh 405
1308 TACCGTTACTTGAAGTAAGCGCGTGGCAACGACCGCTGTCACAAA 1357
405 rAsnValThrTrpLysValAsnGlyValAlaGlyAspAsnLeuHisLysL 422
1358 TCGCAAGGACGCTGTCAGCTTCAAGCAAGGGGNAACCAAGCGCTCG 1407
422 euGlyGluGlyThrLeuThrIleAsnGlyThrGlyValAsnProGlyGly 438
1408 ATCAGCGTGGCGAGGTACAGTCACTTTGGATCAGCAGCAGCAGATAA 1457
439 LeuLysThrGlyAspGlyIleValValLeuAsnGlnAlaAspThrAl 455
1458 AGCAAAACAAAGCTTTAGTGAATCGCTTGTACGCGCAGCGGTA 1507
455 aGlyAsnIleGlnAlaPheSerSerValAsnLeuAlaSerGlyArgProt 472
1508 CGGTGCACTGAATCCGATAATCAGTTCAACCCGACAACTCTATTTC 1557
472 hrValValLeuGlyAspAlaArgGlnValAsnProAspAsnIleSerTrp 488
1558 GGCTTTCCGGCGGAGCTTTGGATTTAAACGGGCACTCGCTTCGTCCA 1607
489 GlyTyrArgGlyGlyLysLeuAspLeuAsnGlyAsnAlaValThrPheTh 505
1608 CCGTATTCAAAATACCGATCAAGGGCGGATGTTGNCNATCATATGCCA 1657
505 rArgLeuGlnAlaAlaAspTyrGlyAlaValIle...ThrAsnAsnAlaG 521
1658 CAACAACATCCCGTTCACCAT...ACAGGGAATGAAAGTATT 1698
521 lnGlnLysSerGlnLeuLeuLeuAspLeuLysAlaGlnAspThrAsnVal 537
1699 ACACAACCGAGTGGTAAGAATATC..... 1722
538 SerGluProThrIleGlyAsnIleSerPropPheGlyGlyThrGlyThrPr 554
1722 1722

554 oglyAsnLeuTyrSerMetIleLeuAsnSerGlnThrArgPheTyrIleL 571
1723AATGACTTAATTAC 1737
571 euLysSerAlaSerTyrGlyAsnThrLeuTrpGlyAsnSerLeuAsnAsp 587
1738 AGCAAAAGAAATTCCTACACAGGTTGGTTGGCGAGAAA.....CATAC 1781
588 ProAlaGlnTrpGluPheValGlyMetAsnLysAsnLysAlaValGlnTh 604
1782 GACCAAA..... 1788
604 rValLysAspArgIleLeuAlaGlyArgAlaLysGlnProValIlePheH 621
1789ACGAACGGCGGCTCAACCTTGTTCACGCGCCCGCAGAA 1830
621 isGlyGlnLeuThrGlyAsnMetAspValAlaIleProGlnValProGly 637
1831 GACCGCACCCNGCTGCTTTCGGCGGGAACAATTTA...AACGGCAACAT 1877
638 GlyArgLysValIlePheAspGlySerValAsnLeuProGluGlyThrIle 654
1878 CAGCAAAACAAACGCAAACTGTTTTCAGCGCAGACCGCACCGCAGC 1927
654 userGlnAspSerGlyThrLeuIlePheGlnGlyHisProValIleHisA 671
1928 CTTACATCATTTAGGAAGCGGTGCTCAAAATGGAAGGTATCCACAA 1977
671 la...SerIleSerGlySerAlaProValSerLeuAsn..... 682
1978 GGAGAAATCGTGTGGCACAACGACTGCATCNACCGCAGCTTTAAAGCGGA 2027
683GlnLysAspTrpGluAsnArgGlnPheThrMetLys 694
2028 AATTTTCATATTTCAGGGCGGCGGTGATTTCCCGCAATGTT... 2073
694 sThrLeuSerLeuLysAspAlaAspPheHisLeuSerArgAsnAlaSerL 711
2074GCCAAAGTGAAGCGGAT...TGNCATTTGAGCAATCAGCCCAA 2115
711 euAsnSerAspIleLysSerAspAsnSerHisIleThrLeuGlySerAsp 727
2116 GCAGTTT...GGTGTGCGCAGCGCA 2138
728 ArgAlaPheValAspLysAsnAspGlyThrGlyAsnTyrValIleProG 744
2139 TCRAAGC.....CATCAATCTGTACAGCTTCGGACTCGACNG 2176
744 uGluGlyThrSerValProAspThrValAsnAspArgSerGlnTyrGluG 761
2177 GT.....CTG 2181
761 lyAsnIleThrLeuAsnHisAsnSerAlaLeuAspIleGlySerArgPhe 777
2182 ACAATTTGTGCGAANAA.....ANCATTACCGCAGATAA 2216
778 ThrGlyGlyIleAspAlaTyrAspSerAlaValSerIleThrSerProAs 794
2217 AGTGATT.....GCTTCATTGCAATAAGA 2239
794 pValLeuLeuThrAlaProGlyAlaPheAlaGlySerSerLeuThrValH 811
2240 CNGACNTNAGCGCANTGTNAGTNCNCCNATNACGNTNTTNAANCTC 2289
811 isAsp..... 812
2290 NCNGGCGTNCNCACTNAANGCAATCTTAGTGCA...AATGCGGATAC 2336
813GlyGlyHisLeuThrAlaLeuAsnGly..... 821
2337 ACGTTATACAGTCACCCAC...AACGCCACCAAAACGGCAACCTTAGCC 2383
822 .LeuPheSerAspGlyHisIleGlnAlaGlyLysAsnGlyLysIleThrL 838

2422TTAAACGCAACNATCGNTTCGGGCAATGCTTCAATTAACT 2465
856 TyrLeuThrAspGlyTyrAspLeuThrGlyAspAsnAlaAlaLeuGluI 872
2466 AAGCAACACCCGCCACAAACGCGCTG..... 2496
872 eThrArgGlyAlaHisAlaSerGlyAspIleHisAlaSerAlaAlaSerT 889
2497ACGTTTCCGACAACGCTAAGGCAACGTA 2526
889 hrValThrIleGlySerAspThrProAlaGluLeuAlaSerAlaGluThr 905
2527 AGCCATTCCGCACCTCAACGGAATGCTCCCTAGCCGATAGGCAGTATT 2576
906 AlaAlaSerAlaPheAlaGly.....SerLeuLe 915
2577 CCATTTCGAAACAGCGCTTTACCGGACAACTCAGCGGCGAGCAAGANA 2626
915 uGluGlyTyrAsnAlaAlaPheAsnGlyAlaIleThrGlyGlyArgAlaA 932
2627 CAGCATTACACTTAAAGACAGCAAGTGGCGCTGCGCTCAGGACGGA 2676
932 sp...ValSerMetHisAsnAlaLeuThrLeuGlyGlyAspSerAla 947
2677 TTAGCAATTTAAACCTTGACAACGCCACCAATTACACTCAATCCGCCCTA 2726
948 IleHisSerLeuThrValArgAsnSerArgIleSer..... 959
2727 TCGCCACGATGCTGCAGCGCGCAACCGCGAGNGTGCAGACACGCCGC 2776
960SerGluGlyAspArgThr..... 965
2777 GCCGCGTTCGCGCGTTCCTTATTCCTGTTACACGCGCAACTTCGGTA 2826
965 965
2827 GAATCCCGTTTCAACAGCTGACGCTAAGCGGCAAAATTGAACNGTCAAG 2876
966PheArgThrLeuThrValAsn...LysLeuAspAlaThrG 978
2877 AACA...TTCCGCTTTATGTCGGAACTCTCGGCTACCGAGCGCAAAAT 2923
978 ySerAspPheValLeuArgThrAspLeu.....LysAsnAlaAspLysI 993
2924 TGAAGCTGGCGAAAGTTCCGAAGNACTTACACCTTGGCGGTCAACAAT 2973
993 leAsnValThrGluLysAlaThrGlySer..... 1002
2974 ACCGCAACGAAACCGCTAAGCCTCGATCAATTGACGTAGTGAAGGAA 3023
1002 1002
3024 ACACAAACACCGCTGCCGAAACCTTAATTTACCTTCGCAAAACGAA 3073
1003 .AspAsn.....SerLeuAsnValSerPhe..... 1010
3074 ACGTCGATGCGCGCGTGGGTTTACCAACTCATCCGCAAGACGCGGAG 3123
1010 1010
3124 TTCGCCCTGCATAATCCGTCGTAAGAAACAGAGCTTTCCGACAAACTCG 3173
1011MetAsnProAlaGlnGlyGlnAlaLeuAsnIleProLeuVa 1025
3174 CNAAGCAAGCCCAAAACAGGCGGAAAGACACAGCCCAAGCCTTG 3223
1025 lThrAlaProAlaGlyThrSerAlaGluMetPheLysAlaGlyThrArg. 1041
3224 ACGCGCTATTTCGGCGCGCGATGCGCGGAAAGACAGAAAGCGTT 3273
1042ValThrGlyPheSerArgValThrProThrLeuHisVal 1054
3274 GCCGAACCGCGCGCGGCGGGAATAATGCGCATTTATGCGCATTCAGGC 3323
1055AspThrSerGlyGlyAsnThrLysTrpIleLeuAspGI 1067
3324 GGAGGAAGAGAAAAAGGGTGCAGCGGATAAGACACAGCGCNTTGGCGA 3373
1067 y.....PheLysAlaGluAlaAspLysAlaAlaAla 1078
3374 AACAGCGGAAGGAAACCGCGCGGNTTACCACCGCTTCCCGCGCGCC 3423
1078 lAlaLysAlaAspSer..... 1082
3424 CGNGCGCGCGCGGATTTGCCGAACCGCAGCGCCCAACCCCACTCA 3473
1082 1082
3474 ACCCAACCGCAGCGGACCTGATNAGCGGTTATGCCAATAGCGGTTGA 3523
1083PheMetAsnAlaGlyTyrL 1089
3524 GTGAATTTTCCGCCACGCTCAACAGCGTTTTCGCCGTACAGGACGAATTG 3573
1089 yAsnPheMetThrGluValAsnAsnLeu..... 1098
3574 GACCGCGTGTTCGCCGAACCGCGCGCAACGCGTTTGGACAAGCNGCAT 3623
1098 1098
3624 CGGNACACCAACACTACCGTTCCGCAAGATTTCCGCGCGCTACCGCCAAC 3673
1099AsnLysArgM 1102
3674 AACCGACCTGCGCCAAATCGGT.....ATG 3699
1102 etGlyAspLeuArgAspThrAsnGlyAspAlaGlyAlaTrpAlaArgIle 1118
3700 CAGAAAAACCTCGGCAGCGCGCGCTCGCATCTCTTTTCGCACAAACG 3749
1119 MetSerGlyAlaGlySerAlaAspGlyTyrSerAspAsnTyrThrHi 1135
3750 GACGGAACACACTTCGAC.....GACGCGATCGGC.... 3780
1135 sValGlnValGlyPheAspLysLysHisGluLeuAspGlyValAspLeuP 1152
3781AATCGCGCGCTTCCCGCGCGC 3804
1152 heThrGlyValThrMetThrTyrThrAspSerSerAlaAspSerHisAla 1168
3805 GCGTTCGCGCAATACGCGATCGCGAGTTCGACATCGGCATCAGCAC 3854
1169 PheSerGlyLysThrLysSerValGlyGly.....GlyLeuTyrAl 1182
3855 GGGCGCGGTTTTCAGCAGCGCANTCTNTCAGACGCGCATCGGAGGCAAAA 3904
1182 aserAlaLeuPheGluSerGlyAlaTyrIleAspLeuIleGly..... 1196
3905 TCGCGCGCGCTGCTGCTATTCAGGCATTCAGGCAGCATACCGCGCGCT 3954
1197LysTyrIleHisHis.....AspAsnAspTyrThrGlyAsn 1208
3955 TTCGCGCGGATTCGCGATCGAACCGTACATCGGCGCAACCGCTATTTCGT 4004
1209 PheAlaSerLeuGlyThrLysHisTyrAsnThrHisSerTrpTyrAlaGI 1225
4005 CCAAAAGCGGATTACCGCTAC..... 4026
1225 yAlaGluThrGlyTyrArgTyrHisLeuThrGluAspThrPheIleGluP 1242
4027GAAACGTCATATCGCCACCCCGCGCTTCGCTTCAACCGN 4068
1242 roGlnAlaGluLeuValTyrGlyAlaValSerGlyLysThrPhe...Arg 1257
4069 TACCGNGCGGCGATTAAAGCAGATTTATTCATTCAAACCGCGCGCAACAT 4118

```

1258 TrpLysaspGly...AspMetAspLeuSerMetLysAsnArgAspPheSe 1273
4119 NTCCATCACNCTTATTNAGCCTGCTCTATACGATGCGCGTTCGGGCA 4168
1273 rProLeuValGlyArgThrGlyValGluLeuGlyLysThrPheSerGlyL 1290
4169 AA.....GTCCGAACACGGCTCAATACCGN..... 4194
1290 yAspTrpSerValThrAlaArgAlaGlyThrSerTrpGlnPheAspLeu 1306
4195 .....GTATTGGCTCAGGATTTCGGCAAAACCGCGAG 4226
1307 LeuAsnAsnGlyGluThrValLeuArgAspAlaSerGlyGluLysArgII 1323
4227 TCGG.....GAATGGGCGTAAACGCCG 4249
1323 eLysGlyGluLysAspSerArgMetLeuPheAsnValGlyMetAsnAlaG 1340
4250 AAATCAA 4257
1340 InileLys 1342

```

seq_name: sp_bacteria:069746

seq_documentation_block:

ID 069746 PRELIMINARY; PRT; 192 AA.

```

AC 069746;
DT 01-AUG-1998 (TRENBLrel. 07, Created)
DT 01-AUG-1998 (TRENBLrel. 07, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE IGA PROTEASE (FRAGMENT).
OS Neisseria lactamica.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=486;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NL327;
RA Sayers J.R., Vitowski S., Read R.;
RT "Strains of Neisseria lactamica carry an Iga1 protease gene homologue
and some express Iga1 protease activity.";
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ001740; CAA04965.1; -.
DR MEROPS; S06.001; -.
KW Protease.
FT NON_TER 1
FT NON_TER 192
SQ SEQUENCE 192 AA; 21385 MW; E4C6FE902FFA08F8 CRC64;

```

alignment_scores:

```

Quality: 1036.00 Length: 192
Ratio: 5.453 Gaps: 0
Percent Similarity: 98.958 Percent Identity: 97.917

```

alignment_block:

US-09-303-518D-651 x 069746 ..

Align seg 1/1 to: 069746 from: 1 to: 192

```

184 GGCAAGTTGTCAGTCGGGGCGAAGATATTGAGTNTACACAAAAAGG 233
1 GlyLysPheAlaValGlyAlaLysAspIleGluValTrpAsnLysLysGI 17
234 GGAGTTGGTCGCAATCAATGACAAAAAGCCCGATGATTGATTTTCTG 283
17 yGluLeuValGlyLysSerMetThrLysAlaPrometileaspPheServ 34
284 TGGTGTCCGTAAACGGCGTGGCGGATTTGGGGCGATCAATATATTGTG 333
34 alValSerArgAsnGlyValAlaAlaLeuValGlyAspGlnTrpIleVal 50
334 AGCGTGGCACATAACGGCGGCTATAACACGTTGATTTTGGTCGGAAGG 383

```

```

|||||
51 SerValAlaHisAsnGlyGlyTyrAsnAsnValAspPheGlyAlaGluGI 67
384 AAGNAATCCCGATCAGCACCGTCTTTTCTTACCAATTTGTAAGAAAGAA 433
|||||
67 ySerAsnProAspGlnHisArgPheSerTyrGlnIleValLysArgAsnA 84
434 ATTATAAGCCTGACAATTACACACCTTACACGGCGATTATCATTATGCGG 483
|||||
84 snTyrLysProAspAsnSerHisProTyrAsnGlyAspTyrHisMetPro 100
484 CGTTTTCGCTAAATTTTTCACAGATCCAGAACCTGTCGAAATGACGAGTGA 533
|||||
101 ArgLeuHisLysPheValThrAspAlaGluProValGluMetThrGlyAs 117
534 CATGAGGGGGAATACCTATTCCGATAAAAGAAATATCCCGAGCGTGCC 583
|||||
117 pMetArgGlyAsnThrTyrSerAspLysGluLysTyrProGluArgValA 134
584 GCATCGGCTCAGGACACCATATTGCGGCTTATGATGATGACAAACACGCG 633
|||||
134 rGileGlySerGlyHisHisTyrTrpArgTyrAspAspLysHisGly 150
634 GATTTATCTACTCGCGCATGTTAATTGCGGCGATGCGGCGATGCGCCATG 683
|||||
151 AspLeuSerTyrSerGlyAlaTrpLeuIleGlyLysAsnThrHisMetGI 167
684 GGGTTGGGAAATAATGCGGTANTAGTTGAGCGGCGATGCGCCATG 733
|||||
167 nGlyTrpGlyAsnAsnGlyValValSerLeuSerGlyAspValArgHisA 184
734 CCAAGGACTATGGCCCTATGCGGATT 759
|||||
184 laAsnAspTyrGlyProMetProIle 192

```

seq_name: sp_bacteria:069747

seq_documentation_block:

ID 069747 PRELIMINARY; PRT; 194 AA.

```

AC 069747;
DT 01-AUG-1998 (TRENBLrel. 07, Created)
DT 01-AUG-1998 (TRENBLrel. 07, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE IGA1 PROTEASE (FRAGMENT).
OS Neisseria lactamica.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=486;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NL3354;
RA Sayers J.R., Vitowski S., Read R.;
RT "Strains of Neisseria lactamica carry an Iga1 protease gene homologue
and some express Iga1 protease activity.";
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ001739; CAA04964.1; -.
DR MEROPS; S06.001; -.
KW Protease.
FT NON_TER 1
FT NON_TER 194
SQ SEQUENCE 194 AA; 21514 MW; D474FB2FC82DA988 CRC64;

```

alignment_scores:

```

Quality: 1028.50 Length: 194
Ratio: 5.385 Gaps: 1
Percent Similarity: 98.454 Percent Identity: 97.423

```

alignment_block:

US-09-303-518D-651 x 069747 ..

Align seg 1/1 to: 069747 from: 1 to: 194

190 TTTGCACTCGGGCGAAGATATTGAGTNTACACAAAAAGGGAGTT 239

```

|||||
1 PheAlaValGlyAlaLysAspIleGluValTyAsnLysLysGlyGluLe 17
240 GGTCGGCAATCATGACAAAAGCCCGATGATGATTTCTCTGGTGT 289
|||||
17 uValGlyLysSerMetThrLysAlaProMetIleAspPheSerValValS 34
290 CGCGTAACGGCGTGGCGCATTTGGTGGCGCATCAATATATTGACGGTG 339
|||||
34 eArGAsnGlyValAlaAlaLeuValGlyAspGlnTyrlleValSerVal 50
340 GCACATAACGGCGCTATACAAAGCTTCAATTTGGTGGCGGAAGGAAGNAA 389
|||||
51 AlaHisAsnGlyGlyTyAsnAsnValAspPheGlyAlaGluGlySerAs 67
390 TCCGATCAGACACCGTTTCTTACCAAAATGTGAAAGAAATAATTATA 439
67 nProAspGlnHisArgPheSerTyGlnIleValLysArgAsnAsnTyrl 84
440 AGCCTGACAAATTCACACCTTACACCGCGATTANCATATGCCGCTTTG 489
|||||
84 ysProAspAsnSerHisProTyArgnGlyAspTyHisMetProArgLeu 100
490 CATAAATTTGTCACAGATGCAGACCTGTCGAAATGACAGTGCATGAG 539
|||||
101 HisLysPheValThrAspAlaGluProValGluMetThrGlyAspMetAr. 117
540 GGGGATACCTATCCGTAAGAATAATATCCGAGCGTGTCCGCGCATCG 589
|||||
117 gGlyAsnThrTySerAspLysGluLysTyProGluArgValArgIleG 134
590 GCTCAGGACACCACTATTGGCGTTAT...GATGATGACAAACACGGCGAT 636
|||||
134 lySerGlyHisHisTyTrpArgTyArgAspAspAspLysHisGlyAsp 150
637 TTATCCTACTCCGGCGCATGTTAATGGCGGCAATACACATATGCAGGG 686
|||||
151 LeuSerTySerGlyAlaTrpLeuIleGlyGlyAsnThrHisMetGlnG 167
687 TTGGGAAATATGCGTANTTGTTCAGCGCGATGTCGCCATGCCA 736
167 tyTrpGlyAsnAsnGlyValValSerLeuSerGlyAspValArgHisAla 184
737 AGCACTATGGCCCTATGCCGATGTCAGGTGCG 768
184 snAspTyGlyProMetProIleAlaGlyAla 194
seq_name: sp_bacteria:Q99QC6

```

```

seq_documentation_block:
ID Q99QC6 PRELIMINARY; PRT; 1364 AA.
AC Q99QC6;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE SEPA, EXTRACELLULAR SERINE PROTEASE OF THE IGA1 PROTEASE FAMILY,
DE SECRETED BY A C-TERMINAL AUTOTRANSORTER DOMAIN (SECRETED
DE PROTEASE).
CN SEPA.
OS Shigella flexneri.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Shigella.
OX NCBI_TaxID=623;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=M90T.
RX MEDLINE=20566792; PubMed=11115111;
RA Buchrieser C., Glaser P., Rusniok C., Medjari H., d'Hauteville H.,
RA Kunst F., Sansonetti P., Parosot C.;
RT "The virulence plasmid pWR100 and the repertoire of proteins secreted
RT by the type III secretion apparatus of Shigella flexneri.";
RL Mol. Microbiol. 38:760-771(2000).
RN [2]

```

```

RP SEQUENCE FROM N.A.
RC PLASMID=VIRULENCE_PWR501;
RA Venkatesan M.M., Goldberg M.B., Rose D.J., Grotbeck E.J., Burland V.,
RA Blattner F.R.;
RT "Complete DNA Sequence and Analysis of the Large Virulence Plasmid of
RT Shigella flexneri.";
RL Infect. Immun. 0:0-0(2001).
DR EMBL; AL391753; CAC05786.1; -
DR EMBL; AF348706; AAK18385.1; -
DR MEROPS; S06.UPW; -
DR InterPro; IPR000710; IGA_S6.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF02395; IGA1; 1.
DR PRINTS; PR00921; IGASERPTASE.
DR SMART; SM00020; TRYPSIN_SPC; 1.
DR PROSITE; PS0240; TRYPSIN_DOM; 1.
KW Hydrolase; Plasmid; Protease; Serine protease.
SQ SEQUENCE 1364 AA; 145937 MW; 71165A82BD098B9E CRC64;

```

```

alignment_scores:
Quality: 1003.00 Length: 1559
Ratio: 1.258 Gaps: 61
Percent Similarity: 51.123 Percent Identity: 25.978
alignment_block:
US-09-303-518D-651 x Q99QC6 ..
Align seg 1/1 to: Q99QC6 from: 1 to: 1364
22 ACACCGAAACACACCGCAAGCCCGGCGCATCGCTTC 71
||| : : : : : |||
27 ThrCysLysSerHisArg. : : : : : ArgLeuSe 35
72 GCCTGCTTACTTAGCCATATGCTGCTGCGCATTTCTCCCAAGCTT 121
||| : : : : : |||
35 rArgArgValIleLeuThrSerValAlaAlaLeuSerLeuSerAlaT 52
122 GG. : : : : : |||
52 rpProAlaLeuSerAlaThrValSerAlaGluIleProTyrllePhe 68
163 CGCGACTTCCCGAAATAAGCAAGTTTGCAGTCGGCGGCGGCAAGATAT 212
||||| : : : : : |||
69 ArgAspPheAlaGluAsnLysGlyGlnPheThrProGlyThrThrAsn 85
213 TGAGGTNTACACAAAAGGGAGGTGTCGGCAATCAATGACAAAAG 262
||||| : : : : : |||
85 eSerIleTyAspLysGlnGlyAsnLeuValGlyLys...LeuAspLysA 101
263 CCGCGATGATTGATTTTCTGCTGCTGCGGTAAAC. : : : : : 297
||||| : : : : : |||
101 laProMetaLaaAspPheSerSerAlaThrIleThrThrGlySerLeuPro 117
298 ...GGCGTGGCGCATTTGGTGGCGCATCAATATATTGTCAGCGTGCACA 344
||||| : : : : : |||
118 ProGlyAspHisThrLeuTySerProGlnTyValValThrAlaLysHI 134
345 TAACGGCGCTATAACACCGTTGATTTTGGTGGCGGAAGNAATCCCG 394
||| : : : : : |||
134 svalSerGlySerAspThrMetSerPheGly. : : : : : 144
395 ATCAGCACCGTTTCTTACCAATATGTAAGAAATAATTATAAGCCT 444
145 .....TyrAlaLysAsnThrTyThrAla 152
445 GACAATTACACCGCTTACAACGGC...GATTANCATATCGCGTTCGA 491
: : : : : |||
153 ValGlyThrAsnAsnAsnSerGlyLeuAspIleLysThrArgArgLeuSe 169
492 TAAATTTGTCACAGATGCAGAACTGTCGAAATGACGAGTCACATGAGG. 540
||| : : : : : |||
169 rLysLeuValThrGluValAlaProAlaGluVal...SerAspIleGlyA 185

```

541GGGAATACCTATTCCGATAAGAAAAATATCCCGAGCGTGTCGCG 595
185 laValSerGlyAlaTyrGlnAlaGlyGlyArgPheThrGluPheTyrArg 201
586 ATCGGCTCAGACACCACTATTGGCGTTATGATGATGACAAACACGCGCA 635
202 LeuGlyGlyMetGlnTyrVallys.....AspLysAsnGlyAs 215
636 TTATCTCTAC.....TCCGGCGCATGGTTAATGGGGCAATACAC 676
215 nArgThrGlnValTyrThrAsnGlyGlyPheLeuValGlyGlyThrValS 232
677 ATATGACAGGTTGGGAATAATATGGC.....GTANTTATGTTGACGCGC 720
232 erAlaLeuAsnSerTyrAsnAsnGlyGlnMetIleThrAlaGlnThrGly 248
721 GATGTG.....CGCATCGCAACGACTATGGCCCTATGCGGATTCAGAGTGC 767
249 AspIlePheAsnProAlaAsn.....GlyProLeuAlaAsnTyrLeuAs 263
768 GCGAGGCGACAGCGGTTCCGCAATGTTTATATGACAAACCAACAATA 817
263 nMetGlyAspSerGlySerProLeuPheAlaTyrAspSerLeuGlnLysL 280
818 AATGGCTGCTCAACGAGTGTTCACAAACCGGCTACCCCTTATTCGCGCAGG 867
280 yStrpValLeuIleGlyValLeuSerSerGlyThrAsnTyrGly..... 294
868 GAAACGGTTTCCAGCTGATACCAAGATTTGCTTACGATGACATTTA 917
295 ...AsnAsnTyrValValThrGlnAspPheLeuGlyGlnGlnProG 310
918 CAGAGCGGATACATACATCGCTCTNTTTTGAACCGCGCAGTACGACATT 967
310 nAsnAspPheAspLysThrIle..... 317
968 TTTCCTTTTACATCCAAAC..... 987
318 ..AlaTyrThrSerGlyGluGlyValLeuGlnTyrLysTyrAspAlaAla 333
988 AACGGTACGGGTACGTAACAGAAACCAACCAAGGTTTCCAAATCCAAA 1037
334 AsnGlyThrGlyThrLeuThrGlnGlyAsn..... 343
1038 GCTTAAAGTACAGACAGTCCGACTGTGTTGACGAATCTTTGAATGAACTG 1087
344ThrThrTyrAspMetHisGlyLysLysGlyAsnAsp.... 355
1088 ATAAAGAACCGATTACGCGCGAGGGGTGTTAATCAGTACCGTCCAAGG 1137
355 355
1138 TTAACACCGGTGAAACCTTTCTTTATCGATTACGCAACGCGCAACT 1187
356 LeuAsnAlaGlyLysAsnLeuLeuPheThrGly...AsnAsnGlyGluVa 371
1188 CATCTTATCAACAACATCAACAGCGCGCGGCTTTGTTATTTTGAAG 1237
371 lValLeuGlnAsnSerValAsnGlnGlyAlaGlyTyrLeuGlnPheAlaG 388
1238 GTGATTTTACGTCTCGCTGAAACACAGCAACGTAAGTGGCAGCGCGGCG 1287
388 lYAspTyrArgValSerAlaLeuAsnGlyGlnThrTyrMetGlyGlyGly 404
1288 GTTATATACAGTAAGACAGTACCTTACTTGAAGTAACGCGGTGGC 1337
405 lIleIleThrAspLysGlyThrHisValLeuLeuTyrGlnValAsnGlyVal 421
1338 AAACGACCGCTGTGCAAAATCGCAAGGACGCTGACGCTTCAAGCCA 1387
421 aGlyAspAsnLeuHisLysThrGlyGluGlyThrLeuThrValAsnGlyT 438

1388 AAGGGAAAAACCAACGCTCGATCAGCGTGGCGGACGGTACAGTCAATTTTG 1437
438 hrGlyValAsnAlaGlyGlyLeuLysValGlyAspGlyThrValIleLeu 454
1438 GATCAGCAGGACAGCAGTAAAGGCAAAACAAAGCCTTTAGTGAATCGG 1487
455 AsnGlnGlnAlaAspAlaAspGlyLysValGlnAlaPheSerSerValGI 471
1488 CTGTGNTCAGCGGACGGGTGCAACTGAATGCCGATCAATCAGTTCA 1537
471 yIleAlaSerGlyArgProThrValValLeuSerAspSerGlnGlnValA 488
1538 ACCCGCAGAACTCTATTTCGGCTTCGCGGCGGACGTTTGGATTAAAC 1587
488 snProAspAsnIleSerTyrGlyArgGlyGlyArgLeuGluLeuAsn 504
1588 GGGCATTCGCTTTCCTCCAGCGTATTCAAATACCGATGAAGGGCGAT 1637
505 GlyAsnAsnLeuThrPheThrArgLeuGlnAlaAlaAspTyrGlyAlaI 521
1638 GATTGNCNATCATATGCCACAACAATCCACCGTTTACCATT..... 1680
521 eIle...ThrAsnAsnSerGluLysLysSerThrValThrLeuAspLeuG 537
1680 1680
537 lnThrLeuLysAlaSerAspIleAsnValProValAsnThrValSerIle 553
1680 1680
554 PheGlyGlyArgGlyAlaProGlyAspLeuTyrTyrAspSerSerThrLy 570
1681A 1681
570 sGlnTyrPheIleLeuLysAlaSerSerTyrSerProPhePheSerAspL 587
1682 CAGGGAATGAAATATTACACACCGAGTGTGAAGATATCAATAGACTT 1731
587 euAsnAsnSerSerValTyrGlnAsnValGlyLysAspHisAsnLysAla 603
1732AATTACAGCAAAAGAAATGTC 1751
604 lIleAspThrValLysGlnGlnLysIleGluAlaSerSerGlnProTyrMe 620
1752 CTACAACGGTTGGTTTGCAGAGAAAGATACGACCAAAACGACGCGCGC 1801
620 tTyrHisGly.....GlnLeuAsnGlyAsnM 629
1802 TCAACCTTGTTTTACCAGCCCGCAGAGAACGACCGCCNGCTGCTTTC 1851
629 etAspValAsnIleProGlnLeuSerGlyLysAspValLeuAlaLeuAsp 645
1852 GCGGGAACAATTTA...AACGGCAACATCACGCAAAACAAACGCGCAACT 1898
646 GlySerValAsnLeuProGluGlySerIleThrLysLysSerGlyThrLe 662
1899 GTTTTTCAGCGGACAGCAGCAGCGCGCTACATCATTCATTAGGAAGCG 1948
662 uIlePheGlnGlyHisProValIleHisAlaGlyThrThrThrSerSerS 679
1949 GGTGTCAAAAATGGAAGGTATCCCAAGAGAGAAATCGTGTGGGCAAC 1998
679 erGlnSer..... 681
1999 GACTGGATCNACCCGCGCTTTAAAGCGGAAAAATTTCCATATTCAGGCGCG 2048
682 AspTyrGlnThrArgGlnPheThrLeuGluLysLeuLysLeuAspAlaAl 698
2049 GCAGCGCGGTGATTTCCCGCAATGTGTCCAAAGTGGGAAGCGGATTCATT 2098
698 aThrPheHisLeuSerArgAsn...GlyLysMetGlnGlyAspIleAsnA 714
2099 TGAGCAATCACGCCCAAGCAGTGTGTGTCGACCGCGCATCAAGCCAT 2148

```

714 laThrAsnGlySerThrValIleLeuGly.....SerSerArg 726
2149 ACAATCTGTACAGTTCGGACTGGACNGGT...CTGACAAATTTGTGTCGA 2195
727 ValPheThrAspArgSerAspGlyThrGlyAsnAlaValSerValG1 743
2196 ANAANCAATTACCGACCATAAAGTGATTGCTTCACTTAACACACACN 2245
743 uGlySerAlaThrAlaThrVal.....GlyAspGlnSerAspT 757
2246 TNACGGCANTGTNAGCTNNCCNATNACGNTNNNTNAAANCTCNCNGGG 2295
757 yrSerGlyAsnValThrLeuGluAsnLysSerSerLeuGlnIle..... 771
2296 CNTGCNNCACTNAANGCAATCTTAGTCAAAATGGCGATACAGTTATAC 2345
772 MetGluArgPheThrGlyIleGluAlaTyRAspSerThrValSerVa 788
2346 AGTCAGCCACACGCCACCAACGCAACCTTAGCTCTGGGCAATG 2395
788 lThrSerGluAsnAlaValPheAspArgValGlySerPheValAsnSerS 805
2396 CCCAGCAACATTTAAATCAAGCCACATTAACGCGCAACNATCGGNTTCG 2445
805 erLeuThrLeuGlyLysGlyAlaLysLeuThrAlaGlnSerGlyIlePhe 821
2446 GCGAATGCTTCATTTAATCTAAGCAACAGCGCGCACAAACGCGAGTCT 2495
822 SerThrGlyAlaValAspValLysGluAsnAla.....SerLe 834
2496 GACGCTTCGACACACGCTAAGGCAACAGTAAGCAT..... 2532
834 uThrLeuThrGlyMetProSerAlaGlnLysGlnGlyTyTySerProV 851
2533 .....TCGCGACTCAACGGCAATGTCCTAGCCGATAAGGCAAGTATTC 2577
851 alIleSerThrThrGluGly...IleAsnLeuGluAspAsnAlaSerPhe 866
2578 CATTTTGAACACGCGCTTACCGGACAACTACGCGCAGCAGGANAC 2627
867 SerValLysAsn.....MetGlyTyRLeuSerSerAsp..... 877
2628 AGCATTTACACTTAAAGACAGCGAATGAGCGCTGCGCTCAGGCACGGAAT 2677
878 .....IleHisAlaGlyThrThrA 884
2678 TAGGCAATTTAACTTGACACGCCACCACTTACACTCAATTCGCGCTAT 2727
884 laAlaThrIleAsnLeuGlyAspSer..... 892
2728 CGCCACGATGTCGAGCGCGCAACCGCGAGNGTGTACAGACGCGCGG 2777
893 .....AspAlaAspAlaGlyLysThrAspSerProLeuPheSerSerLe 907
2778 CGCGCGTTCGCGCGTTCCTTATTA.....TCGTTTACACGCGCAACTT 2821
907 uMetLysGlyTyRAsnAlaValLeuAlaGlySerIleThrGlyAlaGlnS 924
2822 CGGTAGAATCCGTTTCAACACGCTGACGGTAAC..... 2856
924 erThrValAsnMetIleAsnAlaLeuThrPtyrSerAspGlyLysSerGlu 940
2857 ...GCCAAATTAACNGTCAAGGAACATTCGCTTTTATGTCGGAACCTT 2903
941 AlaGlyAlaLeuLysAlaGlySerArgIleGluLeuGlyAspGlyLy 957
2904 CGGTACCGAAGCGCAAAATTAAGCTGGCGGAAGTTCCGAAGNACTT 2953
957 shisPheAlaThrLeuGlnValLysGluLeuSerAlaAspAsnThrThrP 974
2954 ACACCTTGGCGGTCAACATACCGGCAACGACCCGTAAGCCTCGATCAA 3003
794 heLeuMetHisThrAsnAsnSer.....ArgAlaAspGln 985
3004 TTGACGGTAGTGGAA.....GGGAAAGCAACAAACCGCTGTCCGA 3044
986 LeuAsnValThrAspLysLeuSerGlySerAsnAsnSerValLeuValAS 1002
3045 AAACCTTAATTTTCAACCTGCAAAACGACACGTGATCCGCGCGGTGGC 3094
1002 pheLeuAsn..... 1005
3095 GTTACCACTCATCCGCAAGACGGCGAGTTCCGCTCATANTCCGTC 3144
1006 .....LysProAla 1008
3145 AAAGAACAGAGCTTTCGACAACTCGCAAGCAGAGCAAGCAAAACA 3194
1009 SerGluMetSerValThr..... 1014
3195 GCGGAAAAAGACAACGCGCAAGCCTTTCAGCGCTGATTGCGCGCGGC 3244
1015 .....LeuIleThrAlaProL 1020
3245 GCGATGCCGCCCAAGACAGCAAGACGCTTTCGGAACCGCGCGCGCGCA 3294
1020 ysGlySerAspGluLysThrPheThr.....Ala 1029
3295 GCGGGGAAATGTCGCGATTATTCAGCGGAGGAGAGAAAAACGGGT 3344
1030 GlyThrGlnGlnIleGlyPheSerAsnValThrProVal..... 1043
3345 CGAGCGGATAAGACAGCGCNTTTCGCAACAGCGCGGAGGAAACCC 3394
1043 eSerThrGluLysThrAspAlaThrLysTrp.....ValLeuThrG 1058
3395 GCGCGNTTACACGCGCTTCCCGCGCGCGCGCGCGCGCGGATTG 3444
1058 lTyThrGlnThrAlaAspAlaGlyAlaSerLysAlaAlaLysAsp... 1073
3445 CCGCAACCGCAGCGCAACCGCAACCTCAACCCCAACCGCAGCGGACCT 3494
1073 ..... 1073
3495 GATNAGCGCTTATGCCAATAGCGGTTTGTAGTGAATTTCCCGCACGCTCA 3544
1074 .....PheMetAlaSerGlyTyRLeuSerPheLeuThrGluValA 1087
3545 ACAGCGTTTTCGCGGTACAGGAGGATTCGACCGCGTGTTCGCGGAAGAC 3594
1087 snAsnLeuAsnLysArgMetGlyAspLeu.....Argasp 1098
3595 CGCGCGCAAC.....GCNGTTTGGACAGCNGCATCCCGNACACCAACA 3638
1099 ThrGlnGlyAspAlaGlyValTrpAlaArgIleMetAsnGlyThr..... 1113
3639 CTACCGCTTCGAGATTTCGCGCGCTACCGCAACAAACCGACCTCGCGC 3688
1114 ...GlySerAlaAspGlyAspTyRSerAspAsnTyRThrHisValGln. 1128
3689 AAATCGGTATGCAGAAAAACCTTCGCGCGCGGTTCGCGCATCTCTGTTT 3738
1129 IleGlyValAspArgLysHisGluLeuAspGlyValAspLeu..... 1142
3739 TCGCACACCGGACCGCAACANCTTCGACGCGCATCGCAACTCGGC 3788
1143 ...PheThrGlyAlaLeuLeuThrTyRAspSerAsnAlaSerSerH 1158
3789 ACGGCTTCCCGCGCGCTTTCGCGCAATAGCGCATCGCGAGGTTCG 3838
1158 isAla.....PheSerGlyLysAsnLysSerValGly..G 1169
3839 ACATCGCATCAGCAGCGCGCGGTTCGCGCGGTTCGCGCGGTTCGCGCG 3888
1169 ly...GlyLeuTyRAlaSerAlaLeuPheAsnSerGlyAlaTyRPhenAsp 1184
```



```

3889 GGCATCGGAGGCAAAATCCGCCCGGTGCTGCTACGATTCAGGCATTGAGGC 3938
|||||
1185 LeuIleGly.....LysTyrLeuHisAspAsnGln.. 1195
3939 ACATACCGCGCGGTTCGCGGATTCGGCATCGCAACCGTCATACATCGCGG 3988
::: |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1196 ....HisThrAlaAsnPheAlaSerLeuGlyThrLysAspTyrSerSerH 1211
3989 CAACGCGCTATTTCGTCACCAAAAGCGATTACCGCTAC..... 4026
::: |||||:|||||:|||||:|||||:|||||:|||||:|||||:
1211 isSerTrpTyrAlaGlyAlaGluValGlyTyrArgTyrHisLeuThrLys 1227
4027 .....GAAACGTCATATCGCCACCGCCCGG 4052
1228 GluSerTyrValGluProGlnIleGluLeuValTyrGlySerValSerG 1244
4053 TCTTGCGTTCAACCGTATACCGCGCGGATTAAGCGAGATTATTTCATCA 4102
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1244 YLysAlaPheSerTyrPheGluAspArgGlyMetAla.....LeuSerMetL 1259
4103 AACCGCGCGCAACACATNTCCATCACNCCTATTNNAGCCTGTCCTATACC 4152
||:|||||:|||||:|||||:|||||:|||||:|||||:
1259 ysAspLysAspTyrAsnProLeuIleGlyArgThrGlyValAspValGly 1275
4153 GATGCGCGCTCGGC.....AAGTCCGAACACGC..... 4182
||| ||||| |||||:|||||:|||||:|||||:
1276 ArgAlaPheSerGlyAspAspTrpLysIleThrAlaArgAlaGlyLeuG 1292
4183 .....GTCAATACCGCNGTATTGCTCAGGATTTCG 4213
1292 yTyrGlnPheAspLeuLeuAlaAsnGlyGluThrValLeuGlnAspAla 1309
4214 GCAAAACCGCGAGTCGCGAATG..... 4236
::: |||||
1309 ergGlyLysArgPheGluGlyGluLysAspSerArgMetLeuMetThr 1325
4237 ...GGCGTAACCGCGCAATCAAA 4257
|||||:|||||:|||||:|||||:|||||:
1326 ValGlyMetAsnAlaGluIleLys 1333
seq_name: sp_bacteria:Q54165
seq_documentation_block:
ID Q54165 PRELIMINARY; PRT; 1366 AA.
AC Q54165;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE SEPA PROTEIN PRECURSOR.
GN SEPA.
OS Shigella flexneri.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Shigella.
OX NCBI_TaxID=623;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=M90T;
RA MEDLINE=96020667; PubMed=7476198;
RX Benjelloun-Touimi Z., Sansonetti P.J., Parsot C.;
RT "SepA, the major extracellular protein of Shigella flexneri:
RT autonomous secretion and involvement in tissue invasion.";
RL Mol. Microbiol. 17:123-135(1995).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
DR EMBL; Z48219; CAA88252.1; -.
DR MEROPS; S06.0UP; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR000710; Iga_S6.
DR Pfam; PF02395; IGAI; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.

```

```

KW Hydrolase; Serine protease; Signal.
FT SIGNAL 1 56 POTENTIAL.
FT CHAIN 57 1089 MATURE SEPA.
SQ SEQUENCE 1366 AA; 146379 MW; 85E8E05F3EF8A7D8 CRC64;

alignment_scores:
  Quality: 1003.00      Length: 1561
  Ratio: 1.263          Gaps: 60
  Percent Similarity: 50.865  Percent Identity: 25.881

alignment_block:
US-09-303-518d-651 x Q54165

Align seg 1/1 to: Q54165 from: 1 to: 1366

22 ACAACCGAAACACACCGCAAGCCCGAAACCGCGCCGCATCCGCTTCTC 71
|||:|||||:|||||:|||||:|||||:|||||:|||||:
27 ThrCysLysSerHisArg.....ArgLeuSe 35
72 GCCTGCTTACTTAGCCATATGCTGCTGCTTGGCATTTCTCCCAAGCTT 121
|||:|||||:|||||:|||||:|||||:|||||:|||||:
35 rArgArgValIleLeuThrSerValAlaAlaLeuSerLeuSerAlat 52
122 GG.....CGGGCACACACTTATTTCGGCATCACTACCACTACTAT 162
|||:|||||:|||||:|||||:|||||:|||||:|||||:
52 rpProAlaLeuSerAlaThrValSerAlaGluIleProTyrGlnIlePhe 68
163 CGCGACTTTCGCAAAATAAAGCAAGTTTCAGTCGGGGCGCAAGATAT 212
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
69 ArgAspPheAlaGluAsnLysGlyGlnPheThrProGlyThrThrAsnI 85
213 TGAGGTNTACAAACAAAAGGGAGTGGTCGGCAAAATCAATGACAAAAG 262
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
85 eSerIleTyrAspLysGlnGlyAsnLeuValGlyLys...LeuAspLysA 101
263 CCCGATGATTGATTTTCTGTGTGTCGCGTAAC..... 297
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
101 laProMetAlaAspPheSerSerAlaThrIleThrThrGlySerLeuPro 117
298 ...GGCGTGGCGCATTCGTTGGCGGATCAATATATTGAGCGTGGCACA 344
|||:|||||:|||||:|||||:|||||:|||||:|||||:
118 ProGlyAspHisThrLeuTyrSerProGlnTyrValValThrAlaLysH 134
345 TAACGCGCGCTATAACAACGTTGATTGTTGGTGGGAAGGAAGNAATCCG 394
|||:|||||:|||||:|||||:|||||:|||||:|||||:
134 sValSerGlySerAspThrMetSerPheGly..... 144
395 ATCAGCACCGCTTTTCTTACCAATTTGTAAGAAAGAAATATATTAAGCCT 444
|||:|||||:|||||:|||||:|||||:|||||:|||||:
145 .....TyrAlaLysAsnThrTyrThrAla 152
445 GACAAATTCACACCTTACACAGGC...GATTANCATATCGCGGTTTGA 491
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
153 ValGlyThrAsnAsnAsnSerGlyLeuAspIleLysThrArgArgLeuSe 169
492 TAAATTTTCACAGATGCAGAACCTGTCGAAATCAGCAGTGACATGAGG. 540
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
169 rLysLeuValThrGluValAlaProAlaGluVal...SerAspIleGly 185
541 ....GGGAATACCTATTTCGATAAAGAAATATATCCGAGCGGTGTCGCG 585
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
185 laValSerGlyAlaTyrGlnAlaGlyArgPheThrGluPheTyrArg 201
586 ATCGGCTCAGACACCACTATTGCGGTTATGATGATGACAAACACGCGGA 635
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
202 LeuGlyGlyGlyMetGlnTyrValLys.....AspLysAsnGlyAs 215
636 TTTATCTCTAC.....TCCGCGCATGGTTAATTGGCGCAATACAC 676
|||:|||||:|||||:|||||:|||||:|||||:|||||:
215 nArgThrGlnValTyrThrAsnGlyGlyPheLeuValGlyGlyThrVal 232
677 ATATGAGGCTTGGGAAATAATATGCG.....GTANTTAGTTTGGCGGC 720

```

```

232 erAlaLeuAsnSerTyrAsnAsnGlyGlnMetIleThrAlaGlnThrGly 248
    : : : : : | | | | | | | | : : : : : : : : : : : : : : :
721 GATGTCGCCCATGCCAGCACTATGCGCTATGCGCTATGCGGATGCGAGTGGCG 770
    | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
249 AspIlePheLeuIleProProAsnGlyProLeuAlaAsnTyrLeuAsnMe 265
    : : : : : | | | | | : : : : : | | | | | : : : : : | | | | |
771 AGCGCAGCGGTTCGCCATGTTTATTTATGACAAACAAACAAATAAT 820
    | | | | | | | | | | | | | | | | | | | | | | | | | | |
265 tGlyAspSerGlySerProLeuPheAlaTyrAspSerLeuGlnLysLys 282
    | | | | | | | | | | | | | | | | | | | | | | | | | | |
821 GGCTGCTCAACGGAGTTTACAAACCGCTACCTTATTCGCGCAGGAA 870
    | | | | | | | | | | | | | | | | | | | | | | | | | | |
282 rpValLeuIleGlyValLeuSerSerGlyThrAsnTyrGly 295
    : : : : : | | | | | : : : : : | | | | | : : : : : | | | | |
871 AACGGTTTCCAGCTGATACGCAAGATGTTTACGATGACATTCACAG 920
    | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
296 AsnAsnTrpValValThrThrGlnAspPheLeuGlyGlnProGlnAs 312
    : : : : : | | | | | : : : : : | | | | | : : : : : | | | | |
921 AGCGGATACACATACCGTCTTNTTGAACCGCGCAGTAACGACATTTT 970
    : : : : : | | | | | : : : : : | | | | | : : : : : | | | | |
312 nAspPheAspLysThrIle 319
    : : : : : | | | | | : : : : : | | | | | : : : : : | | | | |
971 CCTTTACATCCAAAC 390
    : : : : : | | | | | : : : : : | | | | | : : : : : | | | | |
319 LatYrThrSerGlyGluGlyValLeuGlnTrpLysTyrAspAlaAlaAsn 335
    : : : : : | | | | | : : : : : | | | | | : : : : : | | | | |
991 GGTACGGTACGGTAACGAACCAACCAAGGNTTCCATCCAAAGCT 1040
    | | | | | | | | | | | | | | | | | | | | | | | | | | |
336 GlyThrGlyThrLeuThrGlnGlyAsnThrThrTrpGly 348
    : : : : : | | | | | : : : : : | | | | | : : : : : | | | | |
1041 TAAAGTACAGACGTCCGACTGTTTGAACGAATCTTTGAATGAAACTGATA 1090
    : : : : : | | | | | : : : : : | | | | | : : : : : | | | | |
348 348
1091 AAGAACCAAGTTTACCGCGCAGGGGTGTTATCAGTACCGTCCAAGG 1137
    | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
349 349
1138 1138
356 LeuIleLeuAsnAlaGlyLysAsnLeuLeuPheThrGly 371
    : : : : : | | | | | : : : : : | | | | | : : : : : | | | | |
1182 CAACACTCATCTTATCAACACATCAACCAAGCGCGCGGCGTGTGATT 1231
    | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
371 yGluValValLeuGlnAsnSerValAsnGlnGlyAlaGlyTyrLeuGlnP 388
    : : : : : | | | | | : : : : : | | | | | : : : : : | | | | |
1232 TTGAAGGTGATTTTACGCTCTCGCTGAAACACACGAAACGTCGCAAGC 1281
    | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
388 heAlaGlyAspTyrArgValSerAlaLeuAsnGlyGlnThrTrpMetGly 404
    : : : : : | | | | | : : : : : | | | | | : : : : : | | | | |
1282 GCGGCGTTCATATCAGTGAACACAGTACCGTTACTTGGAAAGTAAACGG 1331
    : : : : : | | | | | : : : : : | | | | | : : : : : | | | | |
405 GlyGlyIleIleThrAspLysGlyThrHisValLeuTrpGlnValAsnG 421
    : : : : : | | | | | : : : : : | | | | | : : : : : | | | | |
1332 CTGCGCAACGACCGCTGTCCAAATCGCAAGGACGCTGCGCTGC 1381
    | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
421 yValAlaGlyAspAsnLeuHisLysThrGlyGluGlyThrLeuThrVal 438
    : : : : : | | | | | : : : : : | | | | | : : : : : | | | | |
1382 AAGCCAAAGGGGAAACCAAGCTCGATCAGCGTGGCGGAGGTACAGTC 1431
    : : : : : | | | | | : : : : : | | | | | : : : : : | | | | |
438 snGlyThrGlyValAsnAlaGlyGlyLeuLysValGlyAspGlyThrVal 454
    : : : : : | | | | | : : : : : | | | | | : : : : : | | | | |
1432 ATTTTGGATCAGCAGGACGATGAAGGCAAAACAAAGCCTTTAGTGA 1481
    | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
455 IleLeuAsnGlnAlaAspArgAspGlyLysValGlnAlaPheSer 471
    : : : : : | | | | | : : : : : | | | | | : : : : : | | | | |
1482 AATCGCTTGTGTCAGCGGCGGCTACGGTGAACCTGAATCCGATATC 1531
    : : : : : | | | | | : : : : : | | | | | : : : : : | | | | |
471 rValGlyIleAlaSerGlyArgProThrValValLeuSerAspSerGln 488
    : : : : : | | | | | : : : : : | | | | | : : : : : | | | | |
1532 AGTTCAACCCGACAACTCTATTTCGGCTTTCGCGGCGGAGCTTGGAT 1581
    | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :

```

2290 NCNGGGGNTCCNACTNNAAGCAATCTTAGTCAAAATGGCGATACAGC 2339
774MetGluArgPheThrGlyIleGluAlaTyrAspSerThrVa 788
2340 TTATACAGTCAGCACCAACCCACCAAAAGCGCAACTTAGCTCGTGG 2389
788 lSerValThrSerGlnAsnAlaValPheAspArgValGlySerPheValA 805
2390 GCAATGCCCAAGCAACATTTAATCAAGCCACATTAACACGCAACNATCG 2439
805 snSerSerLeuThrLeuGlyLysGlyAlaLysLeuThrAlaGlnSerGly 821
2440 GNTCCGGCAATGCTTCATTTAATCAAGCAACCAACGCGCCACAAACCG 2489
822 IlePheSerThrGlyAlaValAspValLysGluAsnAla..... 834
2490 CAGTCTGAGCGCTTCGCAACAGCTAAGCAACAGTAAGCCAT..... 2532
835 .SerLeuThrLeuThrGlyMetProSerAlaGlnLysGlnGlyTyrTyrS 851
2533TCCGCACTCAACGGCAATGCTCTCCCTAGCCGATAGGCA 2571
851 erProValIleSerThrGluGly...IleAsnLeuGluAspAsnAla 866
2572 GTATTCCATTTGAACAGCGCGCTTTACCGGACCAACTCAGCGGCGCAAA 2621
867 SerPheSerValLysAsn.....MetGlyTyrLeuSerSerAsp... 879
2622 GGANACAGCATACACTTAAAGACAGCGAATGGAGCGTCCGCTCAGGCA 2671
880IleHisAlaGlyT 884
2672 CGGAATTAGGCAATTTAAACCTTGACAGCGCCACCATTTACACTCAATCC 2721
884 hrThrAlaAlaThrIleAsnLeuGlyAspSer..... 894
2722 GCCTATCGCCAGATGCTGAGCGCGCAACCGGAGNGTGTTCAGACAC 2771
895AspAlaAspAlaGlyLysThrAspSerProLeuPheSe 907
2772 GCGCGCGCGCTTCGCGCGCTTCCTTATTA.....TCCGTTACACCGC 2815
907 rSerLeuMetLysGlyTyrAsnAlaValLeuArgGlySerIleThrGlyA 924
2816 CAACTTCGGTAGAATCCCGTTTCAACAGCTGACGGTAAAC..... 2856
924 laGlnSerThrValAsnMetIleAsnAlaLeuTrpTyrSerAspGlyLys 940
2857GGCAATTCACNGTCAAGAACATTCGCTTTATGTCGGA 2897
941 SerGluAlaGlyAlaLeuLysAlaLysGlySerArgIleGluLeuGlyAs 957
2898 ACTCTTCGGCTACCGAAGCGCAATTCGAAGCTGCGGAAAGTTCCGAG 2947
957 pGlyLysHisPheAlaThrLeuGlnValLysGluLeuSerAlaAspAsnt 974
2948 GNACTTACACCTTGGCGTCAACAATACCGCAACAGCAACCGTAAGCCTC 2997
974 hrThrPheLeuMetHisThrAsnAsnSer.....ArgAla 985
2998 GATCAATTCAGGTAGTCGAA.....GGAAAGACAAACACCGCT 3038
986 AspGlnLeuAsnValThrAspLysLeuSerGlySerAsnAsnSerValLe 1002
3039 GTCCGAAACCTTAATTTACCTCCCTCAAAACGAAACACGTCGATCCCGCG 3088
1002 uValAspPheLeuAsn..... 1007
3089 CGTGGCGTTACCACTCATCGCAAGACGCGAGTTCCGCTGCATAAT 3138
1008Lys 1008

3139 CCGGTCAAGACACAGAGCTTTCCGACAAACTCGGCAAGGCGAGGCCAA 3188
1009 ProAlaSerGluMetSerValThr..... 1016
3189 AAAACAGCGGGAAGAAACAACGCGCAAGCCTTGACGCGCTGATGGCG 3238
1017LeuIlePheAla 1020
3239 CCGGCGCGATGCGCCCGCAAAAGACAGAAACGCTTGCCGAACCGCCCGG 3288
1020 laProLysGlySerAspGluLysThrPheThr..... 1030
3289 CNGCAGCGCGGAAATGTCGCGCATTTATGACGCGGAGGAGAGAAAAA 3338
1031 ...AlaGlyThrGlnGlnIleGlyPheSerAsnValThrProVal..... 1044
3339 ACGGTGCAGCGCGGATAAAGACAGCGGNTTGGCGAAACAGCGCGAAGCGG 3388
1045IleSerThrGluLysThrAspAspAlaThrLysTrp.....ValL 1058
3389 AAACCGCGCGGNTACACCGCTTCCCGCGCGCGCGCGCGCGCGCGG 3438
1058 euThrGlyTyrGlnThrAlaAspAlaGlyAlaSerLysAlaAlaLys 1074
3439 GATTTGCGCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 3488
1075 Asp..... 1075
3489 CGACCTGATNAGCGGTATGCCAATAGCGTTTGAAGTGAATTTCCGCCA 3538
1076PheMetAlaSerGlyTyrLysSerPheLeuThrG 1087
3539 CGCTCACACGCGTTTCCCGGTACAGGACGAAATTTGGACCGCGTGTTCGC 3588
1087 luValAsnAsnLeuAsnLysArgMetGlyAspLeu..... 1098
3589 GAACCGCGCGCAAC.....GNGTTTGGACAGCGCATCCGGNACAC 3632
1099 ArgAspThrGlnGlyAspAlaGlyValTrpAlaArgIleMetAsnGlyTh 1115
3633 CAAACACTACGTTTCGCAAGATTTCCGCGCTTACCGCGCAACAAACGAC 3682
1115 r.....GlySerAlaAspGlyAspTyrSerAspAsnTyrThrHisV 1129
3683 TCGCCCAATCGGTATGACAAACCTCGGACGCGCGCGCGCGCGCGCATC 3732
1129 alGln.IleGlyValAspArgLysHisGluLeuAspGlyValAspLeu.. 1144
3733 CTGTTTTCGCACAAACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCA 3782
1145PheThrGlyAlaLeuLeuThrTyrThrAspSerAsnAlaS 1158
3783 CTCGCGACGCGTTGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCA 3832
1158 erSerHisAla.....PheSerGlyLysAsnLysSerVal 1169
3833 GGTTCGACATCGGCATCAGCACGCGCGCGCGCGCGCGCGCGCGCGCGCTN 3882
1170 Gly..Gly..GlyLeuTyrAlaSerAlaLeuPheAsnSerGlyAlaTyr 1184
3883 TCAGCGCGCATCGGAGCGCAAAATCCGCGCGCGCGCGCGCGCGCGCGCG 3932
1185 PheAspLeuIleGly.....LysTyrLeuHisHisAspAs 1196
3933 TCAGGACGATACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCA 3982
1196 nGln.....HisThrAlaAsnPheAlaSerLeuGlyThrLysAspTyrS 1211
3983 TCGCGCGCAACGCGCTATTTCCGCGCGCGCGCGCGCGCGCGCGCGCGCTAC 4026
1211 erSerHisSerTrpThrAlaGlyAlaGlyValGlyTyrArgTyrHisLeu 1227
4027GAAACGTCATATCGCCAC 4046

[illegible]

Mon Jul 1 09:26:56 2002

689 rGlyThrSerThrPheArgGlyAsnValAsnMetArgHisSerSerLeuT 706

2326AATGGCATACAGTTATATACAGTCAGCCACAC 2358

706 hrValArgAspHisPheThrGlySerIleThrAlaSerAspSerArgIle 722

2359 GCCACCCCAAAACGGCACTTAGCTCTGGGCAATGCCCAACCAACATT 2408

723 AlaValSerSerGluAsnValArgLeuGluGlyAspSerArgLeuThrSe 739

2409 TAAACAACGCCATTAAACGGCAACATCGNTTCGGCAATGCTTCAT 2458

739 rSerAlaLeuThrVal.....SerAspGlyArgL 750

2459 TTAATCTAAGCAACACGCCGCACAAACAGGCAGCTCTGACGCTTCGAC 2508

750 euHisValLysGlyGlyLeuGluThrAspGlyGlyValThrLeu...Asp 765

2509 AACGCTAAGCAACAGTAAGCACTTCGCACTCAACGGCAATGCTCCCT 2558

766 GlyGlyThrLeuLeuValAspGlyGlySerValArgAsnAspVal... 780

2559 AGCCGATAAGCAGTATTCCATTGTTGAAAC.....AGCCGCT 2596

781TyrGluArgLeuLeuAlaTrpSerGluV 790

2597 TTACCGGCACTCAGCGCAGCAGGACAGGACAGGATTAACACTTAAAGAC 2646

790 alArgGlyGlyLeuAsnGlySer..... 797

2647 AGCGAATGGACGCTGCCGTCAGGCACG..... 2673

798 GlyGluTyrAspPheMetThrGlyAlaAlaGlyLeuLeuArgGlyTyrVa 814

2674GAATTAGCAATTTAAACCTTGACAAACGCCAC.....ATTA 2710

814 lArgGlySerAlaGlyAsnValAsnLeuGlnAsnAlaAlaTrpMetMetT 831

2711 CACTCAATTCGCGCTATCGCCACGATGCTGCAGCGCGCAACCGGCAGN 2760

831 hrGlyAsnSerSerValLysHis..... 838

2761 GTGTACAGACCGCGCGCGCTTCCGCGCTTCCCTATTATTCGCTTAC 2810

839LeuGluSerSerGlySerAlaLeuTyrPheSe 849

2811 ACGGCCAACTTCGTTAGATCCCGTTTCAACACGCTGACGGTAACGGCA 2860

849 rArgPro.....GlyGlyGluPheHisThrLeuThrAlaGlySerM 863

2861 AATTGACNCTCAAGAACATTCGCGTTTATGTCGGAACCTCTTCGGCTAC 2910

863 etAspIleSerAspSerValLeuValMetArgThrAspLeu.....His 877

2911 CGAAGCGCAAAATTGAAGCTGGCGGAAAGTTCGGAAGNACTTACACCTT 2960

878 HisSerAspGlnLeuArgValThrGluSerLeuArgGlyLys..... 891

2961 GCGGTCAACAATACCGGCAACCCGTAAGCCTCGATCAATTGACGG 3010

892AsnAsn.....LeuLeuL 896

3011 TAGTGAAGGGAAGACAAACACCGCTGTCCGAAACCTTATTTCACCC 3060

896 euValAspPheThrGluArgSerAspGlyGlnLysAlaLeuAsnIlePro 912

3061 CTGCAAAACCAACACGCTCGATGCGCGCGCTGCGGTACCACTCATCCG 3110

913 LeuValThrAlaProAlaGlyThrGlyAlaAspValPheSerValLysTh 929

3111 CAAAGCGCGAGTTCCGCTCGCATATCCGGTCAAGAACCAAGACGCTTT 3160

929 rArgAspThrGlyPheSerHisIleThrProValValArgAlaGlu.... 944

336 sphisileileasn..... 340

1069 GAATCTTTGAATCAAACTGATAAAGAACACCGCTTTACGCGGCGAGGGGTGT 1118

341 GlyHisLysAsnAsnProAspTyr..... 349

1119 TAATCAGTACCGTCCAGGTTAAACAACGCGTAAACCTTCTCTTTTATCG 1168

350 GlyAspAsnIleValPhe...S 356

1169 ATTACGGCAACGGCAAACTCATCTTATCAAAACAACATCAACCAAGGCGG 1218

356 erGlyGluGluGlyValIleAsnIleLysAsnAspIleAlaGlnGlyAla 372

1219 GCGGTTTGTATTTTGAAGGTGATTTTACGGTCTCGCTCGCTGAAACAACGA 1268

373 GlyPheIleGlnPheAsnAsnTyrIleValSerThrGlnAsnAsnSe 389

1269 AACGTGTCAAGCGCGCGGCTTCATATCATAGTGAAGACAGTACCGTTACTT 1318

389 rThrIrrThrGlySerGlyIleIleValAsnAspGlySerGluValTyr 406

1319 GGAAAGTAAACGGCGTGGCAACGACCGCTGTCCAAATCGGCAAGGC 1368

406 rpGlyIleAsnGlyValLysAspAspAsnLeuHisLysLeuGlySerGly 422

1369 ACGTGTCAACGTTCAAGCCAAAGGGGAAACCAAGGCTCGATCAGCGTGGG 1418

423 LysLeuIleIleAsnGlyAsnGlyIleAsnThrGlyGlyLeuLysValG 439

1419 CGACGGTACAGTCATTTTGGATCAGCAGCAGCAGATAAAGCAAAAC 1468

439 yAspGlyThrValLeuLeuSerGlnHisLysAspAlaGluGlyArgValG 456

1469 AAGCTTTAGTGAATCGGCTTGCTCAGCGGCGACGGGTACGGTGAACGTG 1518

456 InAlaPheSerSerValAsnIleSerSerGlyArgProValValValLeu 472

1519 AATCGCGATAATCAGTTCAACCCCGCAAACTCTATTTCGGCTTCGCGG 1568

473 SerAspAspLysGlnIleAsnProAspAsnIleSerTrpGlyTyrArgG 489

1569 CGGAGTTTGGATTTAAACGGCATTCGCTTCGTTCCACCGCTATTCAAA 1618

489 yGlyThrLeuAspLeuAsnGlyAsnAsnIleThrPheThrArgIleGlnA 506

1619 ATACGATGAAGGGCGCATGATTCNCNATCATAT 1653

506 IaSerAspTyrGlyAlaIleIleThrAsnAsnAsnAspLysLysSerIle 522

1654GCCACAACAAC 1664

523 ThrAsnLeuAsnIleSerThrLeuLysAspThrAspIleAsnValSerVa 539

1665 ATCCACCGTTACCATACAGG..... 1686

539 lSerAspValHisValLeuGlyGlyLysGlyLysProGlyAspLeuPhe 556

1686 1686

556 yAspArgGlyThrGlyArgTyrLeuIleLeuLysGluLysTyrSer 572

1687AATCAAGTATTACAAACCGAGTGGTAA 1715

573 ProPhePheGlnAspValAsnAsnThrSerValTrpGlnTyrValGlyLys 589

1716 GAATATCATACACTT.....AATT 1735

589 sAsnArgAsnGlnAlaIleAlaValLysSerGluLysIleLysAsnS 606

1736 ACAGCAAAAGAAATTGCCTACACGCTTGGTTTGGCGAGAAAGATACGACC 1785

606 erSerLysProTyrLeuPheHisGly..... 614

60 rOtyrGlnThrPheArgAspPheAlaGluAsnLysGlyValPheThrPro 76

199 GGGCGAAAGATATTGAGTGTNTACAAACAAAGGGGAGTGTGTCGCGCAA 248

77 GlyAlaThrGlyIleGluIleLysAspLysAsnGlyAsnLeuThrGly.. 92

249 ATCAATGACAAAGCCCGATGATTTTTCGTGCTGCTGCGTAACG 298

93 ...ValLeuAsnThrProMetLeuAspPheSerSerLeuSerArgAspG 108

299 GCGTGGCGCATTTGTGGCGCATCAATATATTTGACCGTGGCAATAC 348

108 lYHisThrThrLeuIleHisProGlyTyrGlyValSerAlaLysHis... 123

349 GGGCGGTATACACAGTGTGATTTTGGTGGCGGAAGAGNAATCCGATCA 398

124 GlyArgLeuGlnSerLeuLysThrAlaThrPheGlyTyrAsnAlaGlu.. 139

399 GCACCGTTTTCTTACCAAAATGTGAAAGAAATAATATATAAGCCTGACA 448

140TyrLysIleValAsnAsnHis..... 147

449 ATTACACCGCTTACACGGCGGATTANCATATGCGCGTTTCCATAAATT 498

148AlaSerTyrAspPheThrAlaProArgTyrSerLysLeu 160

499 GTCACAGATGCAGAACTGTCCAAATGACAGTGCACATGAGGGGGAATAC 548

161 ValThrGluValIleProValasn...ThrGlyAsp.....GlyLys 173

549 CTATTCGGATAAGAAATATATCCGAGCGGTGTCGCGATCGGCTCAGGAC 598

173 sTyrAsnAspSerLysLysTyrThrAlaPheTyrArgThrGlyValGlyA 190

599 ACCACTATTGGCCTTATGATGATGACAAACACGCGCAT..... 636

190 spGlnTyrIleLys.....AspLysLysGlyAsnHisThrLeuLeu 203

637TTATCTTACTCCGCGCATGGTTAATTCGCGCAATACATAT 680

204 AsnArgTyrAlaProGlnAsnAlaTyrLeuThrGlyGlyThrValGlyTh 220

681 GCAGGGTTGGGA.....AATAATGGCGTANTTAGTTTACGCGCG 721

220 rProTyrTyrGlyArgValSerAsnGlnThrValLeuIleSerSerProG 237

722 ATGTGCGCCATGCCAGGACTATGCCCTTATGCCGATTCAGTCCGCGCA 771

237 lYAsnThrPheAsnGlnThrHisGlyProLeuAlaSerTyrGlyGlnSer 253

772 GGCACAGCGTTCGCGCAATGTTTATATGACAAACAAACAATAAATG 821

254 GlyAspSerGlySerProLeuTyrAlaTTrpAspAsnIleGlnAsnLysTr 270

822 GCTGCTCAGCGAGTTTACAAACCGGCTACCTTATTCGCGCAGGAAA 871

270 pValLeuAlaGlyValThrLeuTyrAsnGluGlyValAsnGlySerArgA 287

872 ACGGTTTCCAGCTCATACGCAAGATTTGGTTTCCAGATGACATTTACAGA 921

287 snTrpTrpLeuGluLeuProAsnSerPheIleAsnGluThrIleLysSer 303

922 GCGCATACATACCGTCTNTTTTGAACCCCGCAGTAAACGACATTTTC 971

304 AspSerLysGluIleValValLysAsnLysLysGluAspIleThrTrpSe 320

972 CTATTATCCCAACACAGGTGACGGTACGGTAACA...GAACCAACG 1018

320 rPhe...AspLysThrGlnGlyThrGlyValIleAsnLeuGluArgLysA 336

1019 AAAAGGTTTCCAATCCAAGCTTAAGATACAGACAGTCCGACTGTTTGCAC 1068


```

363 lAlaAspGlyLyLysValValTrpGlnValLysAsnProAsnGlyAsp 380
1346 GCGTGTCCAAATCGGCAAGACGCGTGCACGTTCAAGCCCAAGGGAA 1395
380 rgLeuAlaLysIleGlyLysGlyThrLeuGluIleAsnGlyThrGlyVal 396
1396 AACCAAGCTCGATCAGCGTGGGGCAGCGGTACATGTCATTTGGATCAGCA 1445
397 AsnGlnGlyGlnLeuLysValIleGlyAspGlyThrValIleLeuAsnGlnLy 413
1446 GGCACAGCATAAAGGCAAAACAAAGCGTTTGTAGTAATCGGCTTGNTCA 1495
413 sAlaAspSerAsnGlnLysValGlnAlaPheSerGlnValGlyIleValS 430
1496 GCGGCAAGGGGTACGTTGCACTGAATCCGATATCATCTAGTTCAACCCGCAC 1545
430 erGlyArgGlyThrLeuValLeuAsnSerSerAsnGlnIleAsnProAsp 446
1546 AAACCTCTATTTCGGCTTTCGGCGCGGACGTTTGATTTTAAACGGGCAATTC 1595
447 AsnLeuTyrPheGlyPheArgGlyGlyArgLeuAspAlaAsnGlyAsnAs 463
1596 GCTTTCGTTCCACCGTATTCAAATACCGATGAAGGGCGCATGATTCNCN 1645
463 pleuThrPheGluHisIleArgAsnValAspGluGlyAlaArgIleValA 480
1646 ATCATATATGCCAACACATCCACCGTTTACCATTACAGGGAATGAAAGT 1695
480 snHisAsnThrGlyHisAlaSerThrIleThrLeuThrGlyLysSerLeu 496
1696 ATTACACAACCGAGTGGTAAAGATATCAAT 1725
497 IleThrAspProLysThrIleSerIleHis 506

```

alignment block:

lock:

100

alignment block:

```

US-09-303-518D-651 x Q51168
Align seg 1/1 to: Q51168 from: 1 to: 508

202 GCAGAAAGATATTCAGGTNTACACAAAAGGGAGTTGGTGGCGAAATC 251
|||||
1 AlathAspLeuSerValLysAsnLysGlnGlyGlnAsnIleGlyAsnAl 17
252 AATGACAAAAGCCCGATGATTGTTTCTGTGGTGGCGTAAC...G 298
|||||
17 aLeuSerAsnValProMetIleAspPheSerValAlaAspValAsnLysA 34
299 GCGTGGCGGATGTTGGTGGCGATCATATATTTGTGAGCGTGCCACAT... 345
|||||
34 rgIleAlaThrValValAspProGlnTyrAlaValSerValLysHisAla 50
346 .....AACGGCGGTATACACAGCTTGATTT 371
51 LysAlaGluValHisThrPheTyrTyrGlyGlnTyrAsnGlyHisAsnAs 67
372 TGGTGGCGAAGAGNAATCCCGATCAGACCGCTTTTCTTACCAAAATG 421
|||||
67 pValAlaAspLysGluAsn.....GluTyrArgValV 78
422 TGAAAGAAATATTAAGCCTGACAAATTCACACCCCTTACACAGGC... 468
|||||
78 aGluGlnAsnAsnTyrGluProHisLysAlaTrpSerAlaSerAsnLeu 94
469 .....GATTANCAATATCCCGCTGACAGTGCATGAGGGG...AATACCTATT 553
|||||
507 TGCAGAACCTCTCAATGACGAGTGCATGAGGGG...AATACCTATT 553
|||||
111 uValAlaProIleAlaProThrAspAlaGlyGlyLeuAspThrTyrL 128
554 CGGATAAAGAAAATATCCCGAGCGGTGCGCATCGCGTCCAGCACACCAC 603
|||||
128 ysAspLysAsnArgPheSerPheValArgValGlyAlaGlyArgGln 144
604 TAT.....TGCGGTATGATGATGACACAAACACAGGC... 633
145 LeuValTyrGluLysGlyAlaTyrHisGlnGluGlyAsnGluLysGlyTy 161
634 .....GATTTATCTACTCCGCGCATGTTTAAATGGCGGC... 669
161 rAspLeuArgAspLeuSerGlnAlaTyrArgTyrAlaIleAlaGlyThrP 178
670 .....AATACACAT...ATCCAG 684
178 roTyrLysAspIleAsnIleAspGlnThrMetAsnThrGluGlyLeuIle 194
685 GGTGGGGAATAAAGCGGTANTAGTTTGGAGCGCGATGTCGCCATGC 734
195 GlyPheGlyAsnHisAsnThrHisTyrSerAlaGluGluLysGlnAl 211
735 CAACCACTATGCCCCATGCCAGTATGCGAGTGGCGGAGCGACAGCGGTT 784
211 aLeuSerGlnAspAlaLeuThrAsnTyrGlyValLeuGlyAspSerGlyS 228
785 CGCCAATGTTTATTATGACAAAACAAATAATGGCTGCTCAACGGA 834
228 erProLeuPheAlaPheAspLysGlnLysAsnGlnTrpValPheLeuGly 244
835 GTTTTACAAACCGGTACCTTATTCGCGAGGAAACAGGTTTCCAGCT 884
245 ThrTyrAspTyrTrpAlaGlyTyrGly..... 253
885 GATACGCAAGATGTTTCTAGCATGACATTTAC..... 918
254 ....LysLysSerTrpGlnLysTrpAsnIleTyrLysLysGluPheAlaA 269
919 .....AGAGCGGATACACATACCGTCTNTTTTGAACCGCGC 954

```

```

seq_name: sp_bacteria:Q51165
seq_documentation_block:
ID Q51165 PRELIMINARY; PRT; 496 AA.
AC Q51165;
DT 01-NOV-1996 (TRENBLREL. 01, Created)

```



```

408 LysValGlnAlaPheSerGlnValGlyIleValSerGlyArgGlyThrLe 424
1512 GCAACTGAATCCGATATCAAGTTCACCCGACAAACTCTATTTCGGCT 1561
424 uValLeuAsnSerSerAsnGlnIleAsnProAspAsnLeuThrPheGlyP 441
1562 TTCGGCGGACGTTTGGATTAAACGGGCATTCGCTTTCGTCACCGT 1611
441 heArgGlyGlyArgLeuAspAlaAsnGlyAsnAspLeuThrPheGluHis 457
1612 ATTCAAATACCGATGAAGGGCGCATGATGTCNCATCATATATGCCAAC 1661
458 IleArgAsnValAspGluGlyAlaArgIleValAsnHisAsnThrSerHi 474
1662 AACATCCACCGTTACCATACAGGAATGAAGATTATACACACCGAGTG 1711
474 sAlaSerThrIleThrLeuThrGlyLysSerLeuIleThrAsnProAsn 491
1712 GTAAGAATATCAAT 1725
491 erLeuSerValHis 495
seq_name: sp_bacteria:Q51163
seq_documentation_block:
AC Q51163 PRELIMINARY; PRT; 496 AA.
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE IGAI PROTEASE (FRAGMENT).
GN IGAI.
OS Neisseria meningitidis.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=487;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=81139; PubMed=7783620;
RX MEDLINE=95302961;
RA Lonholt H., Poulsen K., Mogens K.;
RT "Comparative characterization of the iga gene encoding IgA1 protease
RT in Neisseria meningitidis, Neisseria gonorrhoeae and Haemophilus
RT influenzae.";
RL Mol. Microbiol. 15:495-506(1995).
DR EMBL: X82477; CAA57860.1; -.
DR MEROPS: S06.001; -.
DR InterPro: IPR000710; IGA_S6.
DR Pfam: PF02395; IGAI; 1.
DR PRINTS: PR00921; IGASERPTASE.
KW protease.
FT NON_TER 1
FT NON_TER 496
FT SEQUENCE 496 AA; 54766 MW; BF4C7A7012FF4D99 CRC64;

alignment_scores:
Quality: 848.00 Length: 538
Ratio: 2.375 Gaps: 17
Percent Similarity: 66.357 Percent Identity: 37.918

alignment_block:
US-09-303-518D-651 x Q51163 ..
Align seg 1/1 to: Q51163 from: 1 to: 496
202 GCGAAAGATATTGAGTNTACACAAAAGGGAGTGTGTCGGCAATC 251
1 AlathrAspLeuSerValLysAsnGlnGlyGlnAsnIleGlyAsnAl 17
252 AATGACAAAAGCCCGATGATTCATTTCTGTGGTGTGCGTAAC...G 298
17 aLeuSerAsnValProMetIleAspPheSerValAlaAspValAsnArg 34
```



```
735 CAACGACTATGGCCCTATCCGATTGCAGTGGCGGAGGAGGAGCGGTT 784
1   :   :   :   :   :   :   :   :   :   :   :   :   :
211 aLeuSerGlnAspAlaLeuThrAsnTyrGlyValLeuGlyAspSerGlyS 228
785 CGCAATGTTTATTATGACAAACAAACAAATAAATGGCTGCCTCAACGGA 834
1111111111111111111111111111111111111111111111111
228 erProLeuPheAlaPheAspLysGlnLysAsnGlnTrpValPheLeuGly 244
835 GTTTTACAAACCGGCTACCGTATATCCGGCAGGAGGAAACCGTTCCAGCT 884
245 ThrTyrAspTyrTrpAlaGlyTyrGlyLysSerTrpGlnGluTrpAs 261
885 CATACGCAAGATGGTTCACGATGACATTTACAGAGGCGGATACACATA 934
1111111111111111111111111111111111111111111111111
261 nileTyrLysLysGluPheAlaAspGluLeuLysGlnArgAsp...AsnA 277
935 CCGTCTNTTGAACCGCGCAGTACAGGACATTTTCCCTTTACATCAAC 984
277 laGlyThrIleLysGlyTyrGlyGluHisTrpLysThrThrGlyThr 293
985 AACACGCTACGGTACGGTAAACAGAACACCAACCAAGAGTNTCCATCC 1034
294 AsnSerHisIleGlySerThr..... 300
1035 AAAGCTTAAAGTACAGACAGTCCGACTGTTTGACGAATCTTTGAATGAAA 1084
301 .....AlaValArgLeuAlaAsn.....AsnGluA 309
1085 CTGATAAAGAACCAAGTTTACCGCGCAGGGGTGTTAATCAGTACCGTCCA 1134
309 tGAsp..... 310
1135 AGGTTAAACACGGTGAAACCTTCTTTTATCGATTACGCAACGGCAA 1184
311 ...AlaAsnAsnGlyGlnAsnValThrPhe.....GluAsnAsnGlyTh 324
1185 ACTCATCTTATCAACACATCAACCAAGCGCGGCGTGTGTTGTTGTTG 1234
324 rLeuValLeuAspGlnAsnIleAsnGlnGlyAlaGlyLeuPhePheL 341
1235 AAGTGATTTTACGTCGCTGCTGCTGAAACACAGCAA....ACGTGGCAAGC 1281
341 ysGlyAspTyrThrValLysGlyAlaAsnAsnAspIleThrTrpLeuGly 357
1282 GCGGGGTTTCATACAGTGAACACATACCGTTACTTGGAAAGTAAACGG 1331
358 AlaGlyIleAspValAlaAspGlyLysLysValValTrpGlnValLysAs 374
1332 CGTGGCAACGACCGCTGTCCAAATCGGCAAGCGACGCTGACAGTTC 1381
374 nProAsnGlyAspArgLeuAlaLysIleGlyLysGlyThrLeuGluIleA 391
1382 AAGCCAAAGGGGAAACCAACGCTCGATCAGCGTGGCGGACGCTACAGTC 1431
391 snGlyThrGlyValAsnGlnGlyGlnLeuLysValIcLyspGlyThrVal 407
1432 ATTTTGCATCAGCAGCAGCAGTAAAGGCAAAACAAAGCCTTTAGTGA 1481
408 IleLeuAsnGlnLysAlaAspAlaAspLysLysValGlnAlaPheSerG 424
1482 AATCGGCTTGNTCAGCGGCGGAGGTACGCTGCACTGATGATCGGATAATC 1531
424 nValGlyIleValSerGlyArgThrLeuValLeuAsnSerSerAsnG 441
1532 AGTTCAACCCGCAACACTCTATTTCCGCTTTCGGCGGCGGAGCTTTGGAT 1581
441 lnIleAsnProAspAsnLeuTyrPheGlyPheArgGlyGlyArgLeuAsp 457
1582 TTAACGGGCAATCGCTTTCGTTCCACCGTATTCATAAATACCGATCAAGG 1631
458 AlaAsnGlyAsnAspLeuThrPheGluHisIleArgAsnValAspGluG 474
1632 GCGCATGATTGNCNATCAATAATGCCACAAACATCCACCGTTTACCATT 1681
```

```
1111111111111111111111111111111111111111111111111
474 yAlaArgIleValAsnHisAsnThrGlyHisAlaSerThrIleThrLeu 491
1682 CAGGGAATCAAGTATTACAAACCGAGTGTGAAGAATATCAAT 1725
1111111111111111111111111111111111111111111111111
491 hrGlyLysLeuLeuThrAspProLysThrIleSerIleHis 505
seq_name: sp_bacteria:Q51164
seq_documentation_block:
ID Q51164 PRELIMINARY; PRT; 508 AA.
AC Q51164;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE IGA1 PROTEASE (FRAGMENT).
GN IGA.
OS Neisseria meningitidis.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=487;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SM1166;
RX MEDLINE=95302961; PubMed=7783620;
RA Iomholt H., Foulson K., Mogens K.;
RT "Comparative characterization of the iga gene encoding Iga1 protease
RT in Neisseria meningitidis, Neisseria gonorrhoeae and Haemophilus
RT influenzae.";
RL Mol. Microbiol. 15:495-506(1995).
DR EMBL; X82486; CAA57869.1; -.
DR MEROPS; S06.001; -.
DR InterPro; IPR000710; IGA_S6.
DR Pfam; PF02395; IGA1; 1.
KW Protease.
FT NON_TER 1
FT NON_TER 508
FT SEQUENCE 508 AA; 56025 MW; FEA91F9BE01B38F CRC64;
SQ
```

alignment_scores:
Quality: 844.00 Length: 560
Ratio: 2.364 Gaps: 19
Percent Similarity: 63.750 Percent Identity: 36.071

alignment_block:

US-09-303-518D-651 x Q51164 ..

Align seg 1/1 to: Q51164 from: 1 to: 508

```
202 GCGAAGATATTGAGTNTACAAACAAAGGGAGTGTGTCGGCAATC 251
1111111111111111111111111111111111111111111111111
1 AlaThrAspLeuSerValLysAsnLysGlnGlyGlnAsnIleGlyAsnAl 17
252 AATGACAAAAGCCCGCATGATTGATTTTCTGTGTCGTCGCGTAAC...G 298
1111111111111111111111111111111111111111111111111
17 aLeuSerAsnValProMetIleAspPheSerValAlaAspValAsnLysA 34
299 CGGTGGCGGCATGTCGGCGCATCAATATATTGTGAGCGTGGCACAT... 345
34 rgIleAlaThrValValAspProGlnTyrAlaValSerValLysHisAla 50
346 .....AACGGCGGCTATAACACAGTGTGATTT 371
51 LysAlaGluValHisThrPheTyrTyrGlyGlnTyrAsnGlyHisAsnAs 67
372 TGGTGGCGGAAGGAGNAATCCCGATCAGCAGCGTGTTCCTTACCAATG 421
1111111111111111111111111111111111111111111111111
67 pValAlaAspLysGluAsn.....GluTyrArgValV 78
422 TGAAGAAGAAATAATTATAAGCTGACAAATTCACACCGCTTACAAAGCGC... 468
1111111111111111111111111111111111111111111111111
78 alGluGlnAsnAsnTyrGluProHisLysAlaTrpSerAlaSerAsnLeu 94
```



```
1471 GCCTTTAGTGAATCGCCTTGNTCAGGGCAGGGTACGGTCAACTGAA 1520
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
410 AlapheserGlnValGlyIleValSerGlyArgThrLeuValLeuAs 426
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1521 TGCGGATAATCAGTTCAACCCGACAAACTCTATTTCGGGTTTCGGCGG 1570
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
426 nSerSerAsnGlnIleAsnProAsnLeuThrPheGlyPheArgGlyG 443
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1571 GACGTTGGATTAAACGGGCATTTCGTTTCACACCGTATTCATAAAT 1620
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
443 lyArgLeuAspAlaAsnGlyAsnAspLeuThrPheGluHisIleArgAsn 459
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1621 ACCGATGAAGGGCGGATGATGNCATCATTAATGCCCACACACATCCAC 1670
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
460 ValAspGluGlyAlaArgIleValAsnHisAsnThrGlyHisThrSerth 476
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1671 CGTTACCATTCACGGGAATGAAGATTAATACACACCGAGTGGTAAGAATA 1720
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
476 rIleThrLeuThrGlySerLeuIleThrAsnProAsnSerLeuSerV 493
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1721 TCAATAGACTT 1731
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
493 alHisSerIle 496

seq_name: sp_bacteria:Q57386
seq_documentation_block:
ID Q57386 PRELIMINARY; PRT; 496 AA.
AC -Q57386;
DT 01-NOV-1996 (T-EMBLrel. 01, Created)
DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (T-EMBLrel. 17, Last annotation update)
DE IGA1 PROTEASE (FRAGMENT).
GN IGA.
OS Neisseria meningitidis.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=487;
RN [1]
RP SEQUENCE FROM N. A.
RC STRAIN=SM1027;
RX MEDLINE=95302961; PubMed=7783620;
RA Lomholt H., Poulsen K., Mogens K.;
RT "Comparative characterization of the iga gene encoding Iga1 protease
RT in Neisseria meningitidis, Neisseria gonorrhoeae and Haemophilus
RT influenzae."
RL Mol. Microbiol. 15:495-505(1995).
DR EMBL; X82472; CAA57855.1; -
DR EMBL; X82470; CAA57853.1; -
DR EMBL; X82471; CAA57854.1; -
DR MEROPS; S06.001; -
DR InterPro; IPR000710; Iga_S6.
DR Pfam; PF02395; IGA1; 1.
DR PRINTS; PR00921; IGASERPTASE.
KW Protease.
FT NON_TER 1 1
FT NON_TER 496 496
SQ SEQUENCE 496 AA; 54634 MW; 4A4F91C0C44AF866 CRC64;

alignment_scores:
Quality: 823.00 Length: 534
Ratio: 2.325 Gaps: 15
Percent Similarity: 66.292 Percent Identity: 36.891

alignment_block:
US-09-303-518D-651 x Q57386 ..
Align seg 1/1 to: Q57386 from: 1 to: 496

202 GCGAAAGATATTGAGTTTACACAAAGGAGTTGGTCGGCAATTC 251
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1 AlaThrAspLeuSerValLysAsnLysGlnGlyGlnAsnIleGlyAsnAl 17
```

```

294 .....|||||:
1124 AGTACCGTCCAAAGTTAAACAACGGTGAACACCTTCTTTATPCGATTAC 1173
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
297 .....GluArgGlyAlaAsnAsnGlyGlnAsnValThrPhe.....Glu 309
1174 GCAACGGCAAACTCATCTATCAACAACATCAACCAAGCGCGGGCGG 1223
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
310 AsnAsnGlyThrLeuValLeuAspGlnAsnIleAsnGlnGlyAlaGlyG 326
1224 TTTGTATTGTAAGGTGATTACGTCTCGCTCGAACAACAAACGAA...A 1270
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
326 yLeuPhePheLysGlyAspTyrThrValLysGlyAlaAsnAsnGlyIle 343
1271 CTTGGCAAGCGCGCGCTCATATCAGTGAAGACAGTACCGTTACTTGG 1320
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
343 hrTrpLeuGlyAlaGlyIleAspValIleAspGlyLysValValTrp 359
1321 AAGTAAACGGCGTGGCAACACCGCTGTCCAAAATCGCAAAAGGCAC 1370
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
360 GlnValLysAsnProAsnGlyAspArgLeuAlaLysIleGlyLysGlyTh 376
1371 GTGCAAGTTCACCAAGCGGAAACCAAGCTCGATCAGCGTGGCG 1420
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
376 rLeuGluIleAsnGlyThrGlyValAsnGlnGlyGlnLeuLysValGly 393
1421 ACGGTACAGTCAATTTGGATCAGCAGCAGCAGATAAAGGCAAAACAA 1470
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
393 spGlyThrValIleLeuAsnGlnAlaAspAlaAspLysLysValGln 409
1471 GCCTTAGTGAATCGGTTGNTACGCGCGAGGTACGGTCAACTGAA 1520
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
410 AlaPheSerGlnValGlyIleValSerGlyArgGlyThrLeuValLeu 426
1521 TCCGATATATCAGTCAACCGGCAAACTCTATTTCGGCTTCGCGCG 1570
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
426 nSerProAspGlnIleAsnProAsnLeuTyrPheGlyPheArgGlyG 443
1571 GACGTTTGGATTAAACGGGATTCGCTTTCCTCCACCGCTATTCAAAAT 1620
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
443 yArgLeuAspAlaAsnGlyAsnAspLeuThrPheGluHileArgAsn 459
1621 ACCGATGAAGCGGCATGATGNCNATCATATGCCACAACATCCAC 1670
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
460 ValAspGluGlyAlaArgIleValAsnHisAsnThrAspArgAlaSerTh 476
1671 CGTTACCATTAAGGAATGAAGTATTACACACCGAGTGGTAAGATA 1720
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
476 rIleThrLeuThrGlyLysSerLeuIleThrAlaProGlnAsnLeuSerV 493
1721 TC 1722
      :|
493 al 493

```

seq_name: sp_bacteria:Q51170

seq_documentation_block:

AC Q51170; PRELIMINARY; PRT; 496 AA.

AC Q51170;

DT 01-NOV-1996 (TrEMBLrel. 01, Created)

DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)

DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)

DE IGAL PROTEASE (FRAGMENT).

GN TGA.

OS Neisseria meningitidis.

OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.

OX NCBI_TaxID=487;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=HF54;

RX MEDLINE=95302961; PubMed=7783620;

RA Lomholt H., Poulsen K., Mogens K.;

```

RT "Comparative characterization of the iga gene encoding IgA1 protease
RT in Neisseria meningitidis, Neisseria gonorrhoeae and Haemophilis
RL Mol. Microbiol. 15:495-506(1995).
DR EMBL; X82473; CAA57856.1; -.
DR MEROPS; S06.001; -.
DR InterPro; IPR000710; Iga_S6.
DR Pfam; PF02395; IGA1; 1.
DR PRINTS; PR00921; IGASERPTASE.
KW Protease.
FT NON_TER 1
FT NON_TER 496
FT NON_TER 496
SQ SEQUENCE 496 AA; 54733 MW; 4BFF91D0D54BB877 CRC64;

```

alignment_scores:

Quality: 815.00 Length: 534
Ratio: 2.309 Gaps: 15

Percent Similarity: 66.105 Percent Identity: 36.704

alignment_block:

US-09-303-518D-651 x Q51170 ..

Align seg 1/1 to: Q51170 from: 1 to: 496

```

202 GCGAAAGATATTGAGTNTACACAAAAAGGGAGTTGTGCGCAATC 251
      ||| |||:|||||:|||||:|||||:|||||:|||||:|||||:
1 AlaThrAspLeuSerValLysAsnLysGlnGlyGlnAsnIleGlyAsnAl 17
252 AATCACAAAAAGCCGATGATTGATTTTCTGTTGGTCTCGCGTAAC...G 298
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
17 aLeuSerAsnValProMetIleAspPheSerValAlaAspValAsnArg 34
299 GCGTGGCGGCATTGGTGGCGCATCAATATATTGTGAGCGTGGCACATAAC 348
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
34 rgThrLeuThrValIleAspProGlnTyrAlaValSerValLysHisVal 50
349 GCGCGCTATAACAACGTTGATTTTGGTGGCGAAGGAGNAATCCCGAT... 396
      ||| |||:|||||:|||||:|||||:|||||:|||||:|||||:
51 LysGlyAspGluIleSerTyrTyrGlyHisHisAsnGlyHisLeuAspVa 67
397 ....CAGCACCGCTTTTCTTACCACAAATTGTGAAAAAGAAATAATTATAAGC 442
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
67 lSerAsnAspGluAsnGluTyrArgSerValAlaGlnAsnAspTyrGluP 84
443 CTGACAAATCA.....CACCTTACACGGC.....GATTANCAT 477
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
84 roAsnLysAsnTrpHisHisGlyAsnGlnGlyArgLeuGluAspTyrAsn 100
478 ATCCCGCGTTTGCATAAATTTGTACAGATGCAGAACCTGTGCAATGAC 527
      ||| |||:|||||:|||||:|||||:|||||:|||||:|||||:
101 MetAlaArgLeuAsnLysPheValThrGluValAlaProIleAlaProTh 117
528 GAGTGACATGAGGGG...AATACCTATTTCGATRAAGAAAAATATCCCG 574
      ||| |||:|||||:|||||:|||||:|||||:|||||:|||||:
117 rSerAlaGlyGlyGlyValGluThrTyrLysAspLysAsnArgPheSerG 134
575 AGCGTGTCCGCTCCGATCAGACACCATATTGGCGTTATCATGATGAC 624
      ||| |||:|||||:|||||:|||||:|||||:|||||:|||||:
134 lUpheValArgValGlyAlaGlyThrGlnPheGluTyrAsnSerArgTyr 150
625 AAACACGGCGATTATCTCTACTCTCCGCGCATGGTTAATTGGCGCAATAC 674
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
151 AsnMetThrGluLeuSerArgAlaTyrArgTyrAlaIleAlaGlyThrPr 167
675 ACATATGACG.....G.....G.....G.....G.....G 685
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
167 oTyrGlnAspValAsnValThrSerAsnLeuAsnGlnGlyLeuIleG 184
686 GTTGGGGAATAATGCGGTANTTAGTTTTCGAGCGCGATGTCGCCCATGCC 735
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
184 lypPheGlyAspAsnSerLysHisHisSerProGluLysLeuLysGluVal 200

```

```
736 AACACATGATGCGCCCTATGCGGATTGCGAGTTCAGAGTTCGCGCAGGCGACACGGTTC 785
    :::::
201 LeuSerGlnAsnAlaLeuThrAsnTyrAlaValLeuGlyAspSerGlySe 217
    :::::
786 GCCAATGTTTATTATGACAAACAAACAAATAAATGCTGCTCAACGGAG 835
    :::::
217 rProLeuPheAlaTyrAspLysGlnGluLysArgTrrpValPheLeuGlyA 234
    :::::
836 TTTTACAA.....ACCGGTACCTTATTCCGCGCAGGGAACGGTTC 879
    :::::
234 laTyrAspTyrTrpAlaGlyTyr.....GlnLysAsnSerTrp 246
    :::::
880 CAG.....CTGATACCAAGATGTTGCTTACGATGACATTTACAGAG 923
    :::::
247 GlnGluTrpAsnIleTyrLysLysGluPheAlaAspGluIleLysGlnAr 263
    :::::
924 CGATACATACATACCTCTNTTTTGAACCGCGCAGTAAACGGACATTTTCCT 973
    :::::
263 gAsp....AsnAlaGlyThrIleLysGlyTyrGlyGluHisTrpLysT 279
    :::::
974 TTACATCAACAACAACGCTACGGGTACGGTACGGTAAACAGAAACCAAG 1023
    :::::
279 hrThrGlyThrAsnSerHisIleGlySerThr..... 289
    :::::
1024 GTNTCCATCAAGCTTAAAGTACAGACGCTCGACTGTTTGTGACGAATC 1073
    :::::
290 .....AlaValargLeu..... 293
    :::::
1074 TTTGAATGAAACTGATAAAGAACCGTTCACGCGCAGGGGTGTTAATC 1123
    :::::
294 .....AlaGlyAsn..... 296
    :::::
1124 AGTACCGTCCAGGTTAAACACGCTGAAACCTTCTTTTATCGATTAC 1173
    :::::
297 .....GluargGlyAlaAsnAsnGlyGlnAsnValThrPhe.....Glu 309
    :::::
1174 GGCACACGCAACTCATCTTATCAACCAACATCAACCAAGCGCGGGCGG 1223
    :::::
310 AsnAsnGlyThrLeuValLeuAspGlnAsnIleAsnGlnGlyAlaGlyG 326
    :::::
1224 TTTGATTTTGAAGTGATTTTACGGTCTCGCTGAAACCAACAGAA...A 1270
    :::::
326 yLeuPhePheLysGlyAspTyrThrValLysGlyAlaAsnAsnGlyIleT 343
    :::::
1271 CCGTGCACAGCGCGCGCTTCATATCAGTGAAGACAGTACCGTACTTGG 1320
    :::::
343 hrTrpLeuGlyAlaGlyIleAspValAlaAspGlyLysValValTrp 359
    :::::
1321 AAAGTAAAGCGGTGGCAACACGCGCTGTCTCAAAATCGCAAGGCAC 1370
    :::::
360 GlnValLysAsnProAsnGlyAspArgLeuAlaLysIleGlyLysGlyTh 376
    :::::
1371 GCTGCAGCTTCAAGCAAGGGAACCAAGGCTCGATCAGCTGGGGC 1420
    :::::
376 rLeuGluIleAsnGlyThrGlyValAsnGlnGlyGlnLeuLysValGlyA 393
    :::::
1421 ACGGTACGCTATTTTGGATCAGCAGGACGCTGCTCAAAATCGCAAGGCAC 1470
    :::::
393 spGlyThrValIleLeuAsnGlnGlnAlaAspAlaAspLysLysValGln 409
    :::::
1471 GCGTTTAGTAAATCGGCTGTTGNTACGCGCAGCGGTACGGTCAACTGAA 1520
    :::::
410 AlaPheSerGlnValGlyIleValSerGlyArgArgThrLeuValLeuAs 426
    :::::
1521 TGCCGATATCAGTCAACCCGACAACTCTATTTCGCTTTCGCGCG 1570
    :::::
426 nSerProAspGlnIleAsnProAsnAsnLeuTyrPheGlyPheArgGlyG 443
    :::::
1571 GAGCTTTGGATTTAAACGGCGCTTCGCTTTCGCTTTCACCGCTATTCAAAT 1620
    :::::
443 lyArgLeuAspAlaAsnGlyAsnAspLeuThrPheGluHisIleArgAsn 459
    :::::
1621 ACCGATGAAGGGCGGATGATGNCNATCATATATGCCACAACATCCAC 1670
```

```
460 ValAspGluGlyAlaArgIleValAsnHisAsnThrAspArgAlaSerTh 476
    :::::
1671 CATTACCATCAGGAATGAAAGTATTACACACCGAGTGGTGAAGAATA 1720
    :::::
476 rIleThrLeuThrGlyLysSerLeuIleThrAlaProGlnAsnLeuSerV 493
    :::::
1721 TC 1722
    ::
493 al 493
seq_name: sp_bacteria:Q48205
seq_documentation_block:
ID Q48205 PRELIMINARY; PRT; 462 AA.
AC Q48205,
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE IGAI PROTEASE (FRAGMENT).
GN IGA.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HK635;
RX MEDLINE=95302961; PubMed=7783620;
RA Lomholt H., Poulsen K., Mogens K.;
RT "Comparative characterization of the iga gene encoding IgA1 protease
in Neisseria meningitidis, Neisseria gonorrhoeae and Haemophilis
influenzae.";
RL Mol. Microbiol. 15:495-506(1995).
DR EMBL; X82488; CAA57871.1; -.
DR MEROPS; S06.001; -.
DR InterPro; IPR000710; IGA_S6.
DR Pfam; PF02395; IGAI; 1.
KW Protease.
FT NON_TER 1
SQ SEQUENCE 462 AA; 51190 MW; 46B634CFF3EE7202 CRC64;

alignment_scores:
Quality: 760.00 Length: 510
Ratio: 2.353 Gaps: 25
Percent Similarity: 63.333 Percent Identity: 38.627

alignment_block:
US-09-303-518D-651 x Q48205 ..
Align seg 1/1 to: Q48205 from: 1 to: 462

202 GCAGAAATATTCAGGTNTACAAACAAAGGGAGTTGTCGCGCAATC 251
    ||| :::::
1 AlaThrAsnValGluValArgAspLysAsnHisSerLeuGlyAsnAl 17

252 AATGACAAAAACCC...CCGATGATGATTTTCTGTGGTCCGCTAAC. 297
    :::::
17 aLeuProAsnGlyIleProMetIleAspPheSerValValAspValAsnL 34

298 ..GCGGTGCGGCATTCGTGGCGATCAATATATTGTAGCGTGGCACAT 345
    :::::
34 ysArgIleGlyThrLeuValAspProGlnIleValSerValLysHis 50

346 AACGCGCGCTATACAAACGTTGATTTT.....GGTGC 377
    ||| ||| ||| |||
51 AlaHisGlnTyrMetAsn...AspPheTyrPheGlyHisTyrAsnGlyHi 66

378 GGAAGAGAGNNAATCCCGATCAGCACCGCTTTTCTTACCAAAATGTGAAA 427
    :::::
66 sArgAspValSerAspAspGluAsnLys.....TyrSerValValThrG 81
```

[illegible]

```

|||||:::|||||||:::|||||:::|||||:::|||||:::|||||:::
334 euLYsAsnSerIleAAsnGInGlyLaGlyGlyLeuPhePheGlUGLyAsn 350
1243 TTTACGGTC...TGCCTCGTAAGAACCAACGAACGTGGCAAAGCGCGCGCCT 1289
   |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
351 TyrThrValGlUGLySerSerAspAsnIleValTrpAsnGLyAlaGLyIL 367
1290 TCATATCAGTGAGACAGTAGTACCCTACTTGGAAAGTAACAACGGCGTGCCAA 1339
   |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
367 eSerValAlaGLyLysThyrValThrTrPLysValHisAsnProGIlnS 384
   |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
1340 ACGACCCCGCTGTCCAATAACGGCAAGGACAGCTGCACGTTCAAGCCAAA 1389
   |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
384 erAspArgLeuAlaLysIlseGLYlYSGlyThrLeuIleValGlUGLyLys 400
1390 GGGGAACAACCAAGCGCTCGATCAGCGTGGGCGACGGTACAGCATATTGTGGA 1439
   |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
401 GlyGUAsnlYsGlySerLeuLysValGLYASPgLYThrVallileLeuLY 417
1440 TCAGCAGGCAGACGATAAAGGCCAAAAAACAAGCCTTTAGTGAATCGGCT 1489
   |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
417 sGInGlnAlaAsPALeAsnAsnLysValLYsAlAPheSeRglNvalIGLYI 434
1490 TGNTACGCGCACGGGTACGCTGCACACTGAATCCGATAATCAGTTC AAC 1539
   |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
434 leValSerGlyArgSerThrValValLeuAsnAspAspLysGlnValasp 450
1540 CCCGACAAACTCTATTTCGGCTTCGGCGC 1569
   |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
451 ProASnSerIleTyRPHeGLYPHeArgGly 460

seq_name: sp_bacteria:O69745

seq_documentation_block:
ID O69745 PRELIMINARY; PRT: 190 AA.
AC O69745:
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE IGAI PROTEASE (FRAGMENT).
OS Neisseria lactamica.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; N
OX NCBI_TaxID=486;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NJ3293;
RA Sayers J.R., Vitowski S., Read R.;
RT "Strains of Neisseria lactamica carry an IgA1 protease gene
and some express IgA1 protease activity.";
RL Submitted (Oct-1997) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AJ001937; CAA04963.1; -.
DR MEROPS; S06.001; -.
KW Protease.
FT NON_TER 1 1
FT NON_TER 190 190
SQ SEQUENCE 190 AA; 21016 MW; 4A524869DB83BAF1 CRC64;

alignment_scores:
    Quality: 733.00      Length: 192
    Ratio: 4.312        Gaps: 1
Percent Similarity: 88.542 Percent Identity: 72.917

alignment_block:
US-09-303-518D-651 x O69745 ..

Align seg i/l1 to: O69745 from: 1 to: 190

205 AAGAATATTGAGTNTACAACAAAAGCGAGTTGGTGC GAATCAAT 254
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
1 LysAspIleGluValTyraSNslYsGlyGluLeuValGLYsSerMe 17

255 GACAAAAGCCCCGATGATTATTTTTTCTGTGGTGTCCGGAACGGCGTGG 304
```



```

2327 ATGGCGATACAGCTTATACAGTCAGCCACACGCCACC...CAAAACGGC 2373
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
721 lylglyileGluAlaTyrAspSerSerValSerValThrSerGlnAsnAla 737
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2374 AACCTTAGCTCGTGGGC.....AATGCCCAAGCAACATTTAAATCA 2414
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
738 ValPheArgValGlySerPheValAsnSerSerLeuThrLeuGluLy 754
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2415 AGCC.....ACATTAAACGGCAACNCATCGGNTTCGGGCAATGCTT 2455
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
754 gLyAlaLysLeuThrAlaGlnGlyIlePheSerThrGly..... 768
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2456 CATTTAATCTAAGCAACACGCCGCGCACAAACGGCAGTCTGCGCTTTC 2505
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
769 .....AlaValAspValLysGluAsnAlaSerLeuLeuThr 781
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2506 GACACGGCTAAGGCAACGTAGCCAT.....TCCGC 2537
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
782 GlyThrProSerAlaGlnLysGlnGlyTyrSerProValIleSerTh 798
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2538 ACTCAACGGCAATGCTCCCTAGCGGATAAGCAGTATTCCATTTTGAA 2587
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
798 rThrGluGly...IleAsnLeuGlyAspLysAlaSerLeuSerValLysA 814
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2588 ACAGCGGCTTTACCGGACAACTCAGC.....GGCAGCAAG 2622
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
814 sn.....MetGlyTyrLeuSerSerAspIleHisAlaGlyThrThr 827
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2623 GANACAGCATTAACCTTA.....AAGACAGCGAATGAGCGTCCGCTC 2666
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
828 AlaAlaThrIleAsnLeuGlyAspGlyAspAlaGluThrAspSerProLe 844
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2667 AGGCACGGAATTA.....GGCAATTAA 2689
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
844 uPheSerSerLeuMetLysGlyTyrAsnAlaValLeuSerGlyAsnIleT 861
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2690 ACCTTGACAGCCACCAATACACTCAATCCGCTATCGCCACGATGCT 2739
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
861 hrglyGluGlnSerThrValAsnMetAsnAlaLeuTrpTyrSerAsp 877
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2740 GCAGCGCGCAAAACGGCGAGTGTCTGACACACGCGCGCGCTTCGCG 2789
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
878 GlyAsnSerThrIleGlyThrLeuLysSerThr..... 888
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2790 CCGTTCCTATTATCCGTTACACGCCCAACTTCGGTAGAATCCCGTTCA 2839
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
888 ..... 888
2840 ACACGCTCACGGTAACGGCAATTTGACNGTCAAGGAACATTCGCGTTT 2889
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
889 .....GlyGlyArgValGluLeuGlyGlyLysAspPhe 900
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2890 ATGTGGAACCTCTTCGGCTACCGAAGCGCAAAATTTGAAGTGGCGGAAG 2939
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
901 AlaThrLeu.....ArgValLysGluLeuAsnAl 910
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2940 TTCGAGAGGNACTTACACCTTGGCGGTCAACAATACCGGCAAGCAACCG 2989
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
910 aAsnAlaThrPheLeuMetHisThrAsnAsnSer..... 922
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2990 TAAGCCTCGATCAATTCACGGTAGTGAAGGGAAGAACACAAACCGCTG 3039
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
923 ..GlnAlaAspGlnLeuAsnValThr.....AsnLysLeuLeu 934
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3040 TCCGAAACCTTTAATTTCAACCTTCGAAACGACACACACACACCGCGC 3089
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
935 GlySerAsn...AsnThrValLeu.....ValAspPhe..... 944
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3090 GTGCGGTTACCACTCATCCGCAAGACGCGAGTTCCGCGCTCATAAATC 3139
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
945 .....LeuAsnLysP 948
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

3140 CGGTCAAAGAAACAGAGCTTTCCGACAAACTCGCAAGGCAAGCCAAA 3189
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
948 roAlaSerGluMetAsnValThr..... 955
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3190 AACACAGCGGGAAGAAAGACAACGCCCAAGCCTTCACGCGGTGATGTCGGC 3239
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
956 .....LeuIleThrAl 959
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3240 CGGCGCGGATCGCGCCGGAAGAACAGACAGAGCGTTGCCGAACCGCGCCGCG 3289
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
959 aProLysGlySerAspGluLysThrPheThr..... 969
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3290 NGGCAGCGGGGAAATGTCCGC.....ATTATGCAGGCGGAGGAA 3330
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
970 ..AlaGlyThrGlnGlnIleGlyPheSerAsnValThrProValIleSer 985
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3331 GAGAAAAAACGGGTGCAGCGGATAAAGACAGACGCGNTTGGCGAAACACGCG 3380
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
986 ThrGluLysThrAspAlaThrLysTrpMetLeuThrGlyTyrGlnTh 1002
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3381 CGAAGCGGAAACCGCGCGGNTACACCGCTTCCCGCGCGCGCGNGCG 3430
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1002 rValSerAspAlaGlyAlaSerLysThrAla..... 1012
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3431 CCGCGCGGATTTCCCGCAACCGCGCCCAACCGCAACCTCAACCCCAA 3480
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1012 ..... 1012
3481 CCGCAGCGGACCTGATNAGCCGTTATGCCAATAGCGGTTTGTAGTGAATT 3530
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1013 .....ThrAspPheMetAlaSerGlyTyrLysSerPh 1023
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3531 TTCGCGCACCTCAACAGCGCTTTCGCGGTACAGGACGAAATTCGACCGCG 3580
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1023 eLeuThrGluValAsnLeuAsnLysArgMetGlyAspLeu..... 1037
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3581 TGTTCGCGAAGACCGCGCAAC.....GCNGTTTGGACACGNCATC 3624
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1038 .....ArgAspThrGlnGlyAspAlaGlyValTrpAlaArgIleMet 1051
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3625 CGGNACACCAACACACTACCGTTTCGCAAGATTTCCGCGCTACGCCCAACA 3674
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1052 AsnGlyThr.....GlySerAlaAsp.GlyGlyTyrSerAspAsnT 1065
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3675 AACCGACTCGCCCAATCGGTATGCAGAAAACCTCGCGACGCGCGCG 3724
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1065 yrThrHisValGlnIleGlyAlaAspArgLysHisGluLeuAspGlyVal 1081
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3725 TCGGCATCTCTGTTTCGCACACCGCGCAACAAACANCTTCGACGACGCG 3774
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1082 AspLeu.....PheThrGlyAlaLeuLeuThrTyrThrAspSe 1094
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3775 ATCGGCACTCGCGACGCTGCCACACGCGCGCTTTCGGGCAATACCG 3824
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1094 rAsnAlaSerSerHisAla.....PheSerGlyLysThrL 1106
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3825 CATCGCAGTTCGCATCGGCATCAGCACGCGCGCGGCTTTTACGACGG 3874
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1106 ysSerValGly..Gly...GlyLeuTyrAlaSerAlaLeuPheAspSerG 1121
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3875 GCANTCTNTCAGACGCGATCGGAGGCAAAATCCGCGCGCGCTGCTGAT 3924
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1121 lYalaTyrPheAspLeuIleGly.....LysTyrLeuHis 1132
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3925 TACGGCATTCAGGCACGATACCGCGCGCTTTCGGCGGATTCGGCATCGA 3974
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1133 HisAspAsnGln.....TyrThrAlaSerPheAlaSerLeuGlyThrLy 1147
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3975 ACCGTACATCGCGCAACGCGCTATTTCGTCCAAAAACGGGATTTACCGCT 4024
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1147 sasPyrSerSerHisSerTrpTyr..... 1155
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
4025 ACGAAACGTCAATATCGCCACCGCGCTTTCGCTTCAACCGNTACCGN 4074
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

1155 ..... 1155
4075 GCGGGCATTAGGCGATTATTCATCAACCGCGCAACACATNTCCAT 4124
||||| ..... |||||
1156 AlaGlyAlaGluValGlyTyrArgTyrHisLeuSerGluSerTrpVa 1172
4125 CACNCCATTATNAGCCTGCTCTATACCGATGCCGCTTCGGGCAAGTC. 4173
||||| ..... |||||
1172 IgluProGlnMetGluLeuValTyr...GlySerValSerGlyLysSerP 1188
4173 ..... 4173
1188 heSerTrpGluAspArgGlyMetAlaLeuSerMetLysAspLysAspTyr 1204
4174 ..... CGAACACGGCTCAATACCCNGTATGCTGCTCAGGA 4208
||||| ..... |||||
1205 AsnProLeuIleGlyArgThrGlyVal.....As 1214
4209 TTTCGGCAAAACCGCAGTGG...GAATGGCGTAAACGCC 4248
||||| ..... |||||
1214 pValGlyArgThrPheSerGlyAspTrpLysIleThrAla 1228

```

seq_name: sp_bacteria:Q48206

```

seq_documentation_block:
ID Q48206 PRELIMINARY; PRT; 492 AA.
AC Q48206;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE IGA1 PROTEASE (FRAGMENT).
GN IGA.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellales;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DA66;
RA MEDLINE=95302961; PubMed=7783620;
RA Lomholt H., Poulsen K., Mogens K.;
RT "Comparative characterization of the iga gene encoding Iga1 protease
RT in Neisseria meningitidis, Neisseria gonorrhoeae and Haemophilus
RT influenzae";
RL Mol. Microbiol. 15:495-506(1995).
DR EMBL; X82467; CAA57850.1; .
DR MEROPS; S06.001; .
DR InterPro; IPR000710; Iga_S6.
DR Pfam; PF02395; IGA1; 1.
DR PRINTS; PR00921; IGASERPTASE.
KW Protease.
FT NON_TER 1 1
FT NON_TER 492 492
SQ SEQUENCE 492 AA; 54751 MW; 6FF93038E7AD5984 CRC64;

```

```

alignment_scores:
Quality: 653.00 Length: 536
Ratio: 2.127 Gaps: 23
Percent Similarity: 57.276 Percent Identity: 32.836

```

```

alignment_block:
US-09-303-518D-651 x Q48206 ..
Align seg 1/1 to: Q48206 from: 1 to: 492

```

```

202 GCGAAGCATATTGAGTNTACACAAAAAGGGAGTGGTGGCGCAATC 251
||||| ..... |||||
1 AlaThrAsnValGluValArgAspLysLysAsnAsnLeuGlySerVa 17
252 AATGACAAAA...GCCCGCATGATTCATTTTCTGTGGTGTGCG...CGTA 295
||||| ..... |||||

```

```

17 lLeuProLysAspIleProMetIleAspPheSerValValAspValAspL 34
296 ACGGGCTGGCGCATTTGGTGGCGGATCAATATTTGGTGGCGTGGCAT 345
:: ..... |||||
34 ysArgIleAlaThrLeuIleAsnProGlnTyrValValGlyValLysHis 50
346 AACGGC...GGCTATAACAACGTTGATTTGGT..... 375
||||| ..... |||||
51 ValGlyAsnGlyValGlyGluLeuHisPheGlyAsnLeuAsnGlyAsnTr 67
376 .... GCGGAAGGAAGNAATCCCGATCAGCACCGT..... 405
||||| ..... |||||
67 pAsnProLysPheGlyAsnSerIleGlnHisArgAspValSerTrpGluG 84
406 ..TTTTCTTTACCAAAATGTGAAAAGAAATAATTAT..... 438
||||| ..... |||||
84 luAsnArgTyrTyrThrValGluLysAsnAsnPheSerSerGluLeuAsn 100
438 ..... 438
101 GlyLysThrGlnAsnAsnGluLysAspLysGlnTyrThrSerAsnLysLy 117
439 ..... AAGCCTGACAATT 451
||||| ..... |||||
117 sAspValProSerGluLeuTyrGlyGlnAlaLeuValLysGluGlnAla 134
452 CACACCTTTACACGGCGATTANCATATATCCGCTTTCATATAATTGTC 501
||||| ..... |||||
134 sGlnLysArgArgGluAspTyrTyrMetProArgLeuAspLysPheVal 150
502 ACAGATGCAGAACCTGTCGAA...ATGACGAGTGACATCAGGGGGAATAC 548
||||| ..... |||||
151 ThrGluValAlaProIleGluAlaSerThrThrSerSerAspAlaGlyTh 167
549 CTATTCGATAAAAGAAATATCCCGAGCGTGTCCGCATCGGCTCA.... 594
||||| ..... |||||
167 rTyrAsnAspGlnAsnLysTyrProAlaPheValArgLeuGlySerGlys 184
595 ..... GGACACCACTATTTGGCGTTATGATGATGAC 624
||||| ..... |||||
184 erGlnPheIleTyrLysLysGlySerHisTyrGluLeuLeuGluGlu 200
625 AACACGCGCATTTATCTACTCCGCGCATGGTTAAATTTGGCGGCAATAC 674
||||| ..... |||||
201 LysAsnGluLysLysGluIleIleHisArgTrpAspValGlyGlyAsp.. 216
675 ACATATCAGCGTGGGGAATAATGCGCTANTTAGTTTTCAGCGGC.... 720
||||| ..... |||||
217 .AsnLeuLysLeuValGlyAsnAlaTyrThrTyrGlyIleAlaGlyThrP 233
721 ..... GATGTGCGCCCATGCCAACGAC.....TATGCC..... 747
||||| ..... |||||
233 roTyrLysValAsnHisThrAspGlyLeuIleGlyPheGlyAspSer 249
748 ..... CCTATGCC 755
||||| ..... |||||
250 ThrGluAspHisAsnAspProLysGluIleLeuSerArgLysProLeuTh 266
756 GATTGCGAGTGGCGGACGCGGATTCAGCTGATACGCAAGATTGG..... 900
||||| ..... |||||
266 rAsnTyrAlaValLeuGlyAspSerGlySerProLeuPheValTyrAspL 283
806 AACACAAACAATAATGGCTCAACGAGTTTTCACAAACCGGCTACCT 855
||||| ..... |||||
283 ysSerLysGluLysTrpLeuPheLeuGly.....AlaTyrAsp 295
856 TATTCGCGGCAAGAAACGGTTTCCAGCTGATACGCAAGATTGG..... 900
||||| ..... |||||
296 PheTrpGly.....GlyTyrLys.....LysLysSerTrpGlnG 307
901 ..... TTCTACGATGACATTTACAGAGCGGATA 928
||||| ..... |||||
307 utrAsnIleTyrLysProGlnPheAlaGluAsnIleLeuLysLysAspS 324

```

```

929 CACATACCGCTCTTTTGAACCGCGAGTAACGACATTTTCTTACA 978
   : : : : : : : : : : : : : : : : : : : : : : : :
324 erAlaGlyLeuLeu.....LysGlyAsnThrGlnTyrAsnTrpThr 337
   : : : : : : : : : : : : : : : : : : : : : : : :
979 TCCAAACAAACAGCGTACGGGTGAGTAACAGAAACCAACAAAGGTC 1028
   : : : : : : : : : : : : : : : : : : : : : : : :
338 SerLysGlyAsn...ThrSerLeuIleSerGlyThrSerGluSerLeuSe 353
   : : : : : : : : : : : : : : : : : : : : : : : :
1029 CAATCCAAAGCTTAAAGTACAGACAGTCCGACTGTTTGTGACGAATCTTGA 1078
   : : : : : : : : : : : : : : : : : : : : : : : :
353 r.....ValAspLeuValAspAsnLys.... 360
   : : : : : : : : : : : : : : : : : : : : : : : :
1079 ATGAAACTGATAAAGAACCAAGCTTTACGGCGGAGGGGTGTTAATCAGTAC 1128
   : : : : : : : : : : : : : : : : : : : : : : : :
360 ..... 360
1129 CGTCCAAAGCTTAAACACCGTGAACACCTTTCTTTTATCAGTACGGCAA 1178
   : : : : : : : : : : : : : : : : : : : : : : : :
361 .....AsnLeuAsnHisGlyLysAsnValThrPhe.....GluGlySe 373
   : : : : : : : : : : : : : : : : : : : : : : : :
1179 CGSAAACTCATCTTATCAAAACATCAACCAAGCGCGGCGGCTTGT 1228
   : : : : : : : : : : : : : : : : : : : : : : : :
373 rGlyAsnLeuThrLeuAsnAsnIleAspGlnGlyAlaGlyGlyLeuP 390
   : : : : : : : : : : : : : : : : : : : : : : : :
1229 ATTTGAAAGTGATTTTACGGTC...TCGCCTGAAACCAACCAACAGCGTGG 1275
   : : : : : : : : : : : : : : : : : : : : : : : :
390 hePheGlyAspTyrGluValLysGlyThrSerGluAsnThrThrTrp 406
   : : : : : : : : : : : : : : : : : : : : : : : :
1276 CAAGCGCGGGCTTCATATCAGTCAACACAGTACCGTTACTTGTGAAAGT 1325
   : : : : : : : : : : : : : : : : : : : : : : : :
407 LysGlyAlaGlyIleSerValAlaGlyLysThrValLysTrpLysVa 423
   : : : : : : : : : : : : : : : : : : : : : : : :
1326 AAACGCGGTGGCAACACGCGCTGTCCAAATCGGCAAGGACGCGTGC 1375
   : : : : : : : : : : : : : : : : : : : : : : : :
423 lHisAsnProGlnPheAspArgLeuAlaLysIleGlyLysGlyLysLeuI 440
   : : : : : : : : : : : : : : : : : : : : : : : :
1376 AGTTTCAGCCAAAGGGGAAACCAAGGCTCGATCAGCGTGGCGACGGT 1425
   : : : : : : : : : : : : : : : : : : : : : : : :
440 leValGluGlyArgGlyAspAsnLysGlySerLeuLysValGlyAspGly 456
   : : : : : : : : : : : : : : : : : : : : : : : :
1426 ACAGTCATTTTGCATCAGCAGCAGCAGCAGTAAAGCAACCAACAGCCTT 1475
   : : : : : : : : : : : : : : : : : : : : : : : :
457 ThrValValLeuLysGlnGlnThrThr.....GlyGlnHisAlaPhe 471
   : : : : : : : : : : : : : : : : : : : : : : : :
1476 TAGTGAATCGGCTTGNTCAGCGGAGGGTACGTCGCAACTGAATGCGC 1525
   : : : : : : : : : : : : : : : : : : : : : : : :
471 eAlaSerValGlyIleValSerGlyArgSerThrValValLeuAsnAspA 488
   : : : : : : : : : : : : : : : : : : : : : : : :
1526 ATAATCAG 1533
   : : : : : : : : : : : : : : : : : : : : : : : :
488 sPasnGln 490

```

seq_name: sp_bacteria:Q48159

```

seq_documentation_block:
ID   Q48159  PRELIMINARY;      PRT;   460 AA.
AC   Q48159;
DT   01-NOV-1996 (TREMBlrel. 01, Created)
DT   01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT   01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE   IGA1 PROTEASE (FRAGMENT).
GN   IGA.
OS   Haemophilus influenzae.
OC   Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC   Haemophilus.
OX   NCBI_TaxID=727;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=287/86;
RX   MEDLINE=96029734; PubMed=7591075;
RA   Lomholt H., Kilian M.;
RT   "Distinct antigenic and genetic properties of the immunoglobulin A1

```

RT protease produced by Haemophilus influenzae biogroup aegyptius associated with Brazilian purpuric fever in Brazil.*;
 RL Infect. Immun. 63:4389-4394(1995).
 DR EMBL; X86103; CAA60056.1; -.
 DR MEROPS; S06.001; -.
 DR InterPro; IPR000710; Iga_S6.
 DR Pfam; PF02395; IGA1; 1.
 KW Protease.
 FT NON_TER 1 1
 FT NON_TER 460 460
 SQ SEQUENCE 460 AA; 51240 MW; 0CCC95B956562F9B CRC64;

alignment_scores:

Quality: 644.00 Length: 500
 Ratio: 2.132 Gaps: 23
 Percent Similarity: 60.400 Percent Identity: 35.600

alignment_block:

US-09-303-518D-651 x Q48159 ..

Align seg 1/1 to: Q48159 from: 1 to: 460

```

202 GCGAAAGATATTGAGTNTACAAACAAAGGGAGTTGTCGCAAAATC 251
   ||| : : : : : : : : : : : : : : : : : : : : : : : :
1 AlathrAsnValGluValArgAspLysAsnHisSerLeuGlyAsnAl 17
   : : : : : : : : : : : : : : : : : : : : : : : :
252 AATGACAAAAGCC...CCGATGATGATTTTCTGTGTCGCGTAAC. 297
   : : : : : : : : : : : : : : : : : : : : : : : :
17 aLeuProAsnGlyIleProMetIleAspPheSerValValAspValAsnL 34
   : : : : : : : : : : : : : : : : : : : : : : : :
298 ..GCGCTGGCGGCATTTGTCGGCGCATCAATATATTGTGACGTCGCACAT 345
   : : : : : : : : : : : : : : : : : : : : : : : :
34 ysArgIleGlyThrLeuValAspProGlnTyrIleValSerValLysHis 50
   : : : : : : : : : : : : : : : : : : : : : : : :
346 AACGGCGCTATAACAAACGTTGATTTT.....GGTGC 377
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
51 AlahisGlnTyrMetAsn...AspPheTyrPheGlyHisTyrAsnGlyHI 66
   : : : : : : : : : : : : : : : : : : : : : : : :
378 GGAAGGAAGNAATCCCGATCAGCAGCGTTTCTTACCAAAATTTGTGAAA 427
   : : : : : : : : : : : : : : : : : : : : : : : :
66 sArgspValSerAsnAspGluAsnLys.....TyrSerValValThrG 81
   : : : : : : : : : : : : : : : : : : : : : : : :
428 GAAATAATTATTAAGCCTGACAAATTCACACCTTACAAAC.....GGC 468
   : : ||||| : : : : : : : : : : : : : : : : : : :
81 lAsnAsnValAsnSerSerGluLysTrpAspValAsnLysArgLeuAsp 97
   : : : : : : : : : : : : : : : : : : : : : : : :
469 GATTANCATATCCGCGCTTTCGATAAATTTGTCACAGATGCAGAACCTGT 518
   ||| : : ||||| : : : : : : : : : : : : : : : : : : :
98 AspTyrAsnMetProArgLeuAsnLysPheValThrGluValAlaProth 114
   : : : : : : : : : : : : : : : : : : : : : : : :
519 CGAAATCAGCAGTGACATGAGGGGGAAT.....ACCTATCCGATA 559
   : : ||||| : : : : : : : : : : : : : : : : : : :
114 r.....ThrProThrLeuAlaGlyAspAspLeuGluThrTyrLysAspL 129
   : : : : : : : : : : : : : : : : : : : : : : : :
560 AAGAAAATATCCGAGCGTTCGCGATCGGTCAGGACACCACTAT... 606
   ||||| : : : : : : : : : : : : : : : : : : : : : : : :
129 ysGluLysTyrProSerPheValArgValGlyAlaGlyArgGlnLeuVal 145
   : : : : : : : : : : : : : : : : : : : : : : : :
607 .....TGGCGTTATGATGATGACAAACACGCG..... 633
   : : : : : : : : : : : : : : : : : : : : : : : :
146 TyrGluLysGlySerArgHisValGluGlyAsnGluHisGlyGluAspLe 162
   : : : : : : : : : : : : : : : : : : : : : : : :
634 ....GATTATTCCTACTCCGCGCATGGTTAATTCGCGCAATACACATA 679
   ||||| : : : : : : : : : : : : : : : : : : : : : : : :
162 ulysAspLeuSerValAlaTyrAsnTyrAlaIleGlyGly...ThrProt 178
   : : : : : : : : : : : : : : : : : : : : : : : :
680 TGCAGGGTTGGGAAATAATGCGGTANTTGTGAGCGC..... 720
   : : ||| : : : : : : : : : : : : : : : : : : : : : :
178 yrGluGlyIleAsnIleAspProSerGlnSerLysGlyLeuIleGly 194
   : : : : : : : : : : : : : : : : : : : : : : : :
721 .....GATGTGCGCCATGCCAAGCA 740
   : : : : : : : : : : : : : : : : : : : : : : : :

```


702 TyrPheGlyAsnValThrLeuAsnAsnHisSerLeuLeuAspIle..... 716
2295 GNTCCNNCACTNAAGCGCAATCTTAGTGCAAAATGGCGATACAGTTATA 2344
717GlyAsnLysPheThrGlyGlyIleGluAlaIle 728
2345 GAGTCAGCACACGCCACC...CAAAACGGCAACCTTAGCTCGTGGC 2391
728 sSerSerValSerValThrSerGlnAsnAlaValPheAspArgValGly 744
2392AATGCCCAAGCAACATTTAATCAAGCC.....ACATT 2423
745 SerPheValAsnSerSerLeuThrLeuGluLysGlyAlaLysLeuThrAl 761
2424 AAACGGCAACNCATCGGTTCCGGCAATGCTTCATTTAATCTAAGCAACA 2473
761 gInGlyGlyIlePheSerThrGly.....AlaValA 772
2474 AGCGCGCACAAAACGCGCTGACGCTTCGCAACGCTAAGGCAAC 2523
772 sValLysGluAsnAlaSerLeuIleLeuThrGlyThrProSerAlaGln 788
2524 GTAAGCCAT.....TCGCACTCAACGGCAATGCTC 2555
789 LysGlnGluTyrTyrSerProValIleSerThrThrGluGly...IleAs 804
2556 GCTAGCCGATAAGCGAGTATTCATTTCAAAACAGCCGCTTACCGGAC 2605
804 nLeuGlyAspLysAlaSerLeuSerValLysAsn.....MetGlyT 818
2606 AACTCAGC.....GGCAAGCAAGANACAGCAATACACTTA 2640
818 yrLeuSerSerAspIleHisAlaGlyThrThrAlaAlaThrIleAsnLeu 834
2641AAGACAGCAATGAGCGTGGCTCAGCAGCGAATTA..... 2679
835 GlyAspGlyAspAlaGluThrAspSerProLeuPheSerSerLeuIleAr 851
2680GGCAATTTAAACCTTGACACGCCACCA 2707
851 gGlyTyrAsnAlaValLeuSerGlyAsnIleThrGlyGluGlnSerThrV 868
2708 TTACTCAATTCGCTATCCGCTATCCGACGATGCTGACGGCGCAACCGCGC 2757
868 aAsnMetAsnAlaLeuTyrTyrSerAspGlyAsnSerThrIleGly 884
2758 AGNGTGTACAGACGCGCGCGCTTCGCGCGCTTCCTATTATCCGT 2807
885 ThrLeuLysSerThrGlyGlyArg..... 892
2808 TACACCGCAACTTCGGTAGAA.....TCCCGTTTCAACACGC 2845
893ValGluLeuGlyGlyLysAspPheAlaThrL 904
2846 TGACGTAACGGCAATTAACNGTCAAGAACATTCGCTTTATGTCG 2895
904 euArgValLys...GluLeuAsnAlaAsnAla...ThrPheLeuMet 918
2896 GAACCTCTCGCTACCAAGCGCAAAATTAAGCTGGCGGAAGTTCCGA 2945
919 HisThrAsnThrSerGlnAlaAspGlnLeuAsnValThrAsnLysLeuLe 935
2946 AGNACTTACACCTTGGCGGTCAACAATACCGCAACGCGGTAAAGC 2995
935 uGlySer.....AsnAsnThr.....ValLeuV 943
2996 TCGATCAATTGACGGTAGTGAAGGGAAGACACAAACCGCTGTCGAA 3045
943 alAspPheLeu.....AsnLysProAlaSerGlu 952
3046 AACCTTAATTACCTGCAAAACGACACGCTGATCGCGCGCGTGGCG 3095
953 ...MetAsnValThr..... 956

3096 TTACCAACTCATCCGCAAAAGACGGGAGTTCCGCCCTGCATATCCGGTCA 3145
956 956
3146 AAGAACAAGAGCTTTCGACAAACTCGGCAAGCGCAGAACCAAAACAG 3195
956 956
3196 GCGGAAAAAGACAAACGCAAAAGCCTTACGCGCTGATTGCGCGCGCGG 3245
957LeuIleThrAlaProly 962
3246 CGATGCCGCGCAAAAGACAGAAAGCCTTGCCGAACCGCGCGCGCGGCG 3295
962 sGlySerAspGluLysThrPheThr.....Alag 972
3296 GCGGGGAAATGTCGCG.....ATTATGACGGGAGGAGAGAAA 3336
972 lyThrGlnGlnIleGlyPheSerAsnValThrProValIleSerThrGlu 988
3337 AAACGGGTGACGGCGGATAAAGACAGCGCNTTGGCGAAACAGCGCGAAGC 3386
989 LysThrAspAlaThrLysTrpMetLeuThrGlyTyrGlnThrValSe 1005
3387 GGAACCGCGCGGNTACCAACCGCTTCCCGCGCGCGCGCGCGCGCGC 3436
1005 rAspAlaGlyAlaSerLysThrAla..... 1013
3437 GGGATTTCCGCAACCGCAGCCCAACCGCAACCTCAACCCCAACCGCAG 3486
1013 1013
3487 CGCAGCTGATNAGCGCTTATGCCAATAGCGTTTGTAGTGAATTTCCCGC 3536
1014ThrAspPheMetAlaSerGlyTyrLysSerPheLeuTh 1026
3537 CACCTCAACAGCGCTTTCGCGGTACAGGACGAAATTCGCGCGCTGTG 3586
1026 rGluValAsnAsnLeuAsnLysArgMetGlyAspLeu..... 1038
3587 CCGAAGACCGCGCGCAAC.....GCNGTTTGGACAGCNGCATCCCGNAC 3630
1039 ..ArgAspThrGlnGlyAspAlaGlyValTrpAlaArgIleMetAsnGly 1054
3631 ACCAAACACTACCGTTCGCAAGATTCGCGCGCTACCGCGCAACAAACCGA 3680
1055 Thr.....GlySerAlaAsp.GlyGlyTyrSerAspAsnTyrThrH 1068
3681 CTGCGGCAATCGGTATGCAGAAAAACCTCGCGAGCGCGCGTCCGCA 3730
1068 isValGlnIleGlyAlaAspArgLysHisGluLeuAspGlyValAspLeu 1084
3731 TCCTGTTTTCGCAACCGCGGCAAAACANCTTCGACGAGCGCATCGGC 3780
1085PheThrGlyAlaLeuLeuThrThrThrAspSerAsnAl 1097
3781 AACTCGGCGCGCTTCCCGCGCGCTTTCGCGCAATACGCGCATCGG 3830
1097 aSerSerHisAla.....PheSerGlyLysThrLysSerV 1109
3831 CAGGTTCGACATCGCATCAGCAGCGCGCGGTTTTCAGCAGCGCGCANTC 3880
1109 aiGly..Gly...GlyLeuTyrAlaSerAlaLeuPheAspSerGlyAlaT 1124
3881 TMTACAGCGCATCGGAGGCAAAATCCCGCGCGCGCTGCTCATTCAGGC 3930
1124 yrPheAspLeuIleGly.....LysTyrLeuHisHisAsp 1135
3931 ATTCAGGCGCATACCGCGCGGTTTCGCGGATTCGCGCATCGAACCGTA 3980
1136 AsnGln.....TyrThrAlaSerPheAlaSerLeuGlyThrLysAspTy 1150

```
3981 CATCGCGCAACGCGCTATTTCGTCCAAAAAGCGGATTACCGCTACGAA 4030
|   :::: |||
1150 rSerSerHisSerTrpTyr..... 1156
4031 ACGTCAATATCGCCACCCCGGCTTGGGTTCAACCGNTACCGNGCGGCG 4080
|   :::: |||
1157 .....AlaGly 1158
4081 ATTAAGGCAGATTATTCATTCAAACCGCGCACACATNTCCATCACNCC 4130
|   :::: ||| :::: :::: :::: ||
1159 AlaGluValGlyTyrArgTyrHisLeuSerGluGluSerTrpValGluPr 1175
4131 TTATTNAGCCTCTCTATACCGATCCGCTTCGGGCAAAGTC..... 4173
|   :::: ||| :::: |||
1175 oGlnMetGluLeuValTyr...GlySerValSerGlyLysSerPheIleT 1191
4173 ..... 4173
1191 rpGluAspArgGlyMetAlaLeuSerMetLysAspLysAspTyrAsnPro 1207
4174 .....CGAACACGCGCTCAATACCGCNGTATTGGCTCAGGATTTCGG 4214
|   :::: |||
1208 LeuIleGlyArgThrGlyVal.....AspValG1 1217
4215 CAAAACCCGCGAGTCG...GAATGGGCGCTAAACGCC 4248
|   :::: ||| :::: |||
1217 yArgThrPheSerGlyAspAspTrpLysIleThrAla 1229
```

